Synthetic biology approaches for the production of chiral aminoalcohols in engineered *E. coli* strains

by

Panwajee Payongsri

A thesis submitted for the degree of Doctor of Philosophy of University College London

Department of Biochemical Engineering

University College London

2014

Declaration

'I, Panwajee Payongsri, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.'

Candidate's signature

.....

Panwajee Payongsri

Acknowledgement

Completing my PhD at the Biochemical Engineering Department of UCL has provided me not only with the research skills but also numerous invaluable experiences. The research would never have been successfully accomplished without a number of people around me.

First of all I would like to thank my supervisor Professor Paul Dalby for his consistent support, dedication and patience. He has given me so much freedom to explore my curiosities and allowed me to freely share my research ideas. This approach led to very useful discussions, rapid progress with my research and a truly enjoyable project. I would like to further extend my gratitude for his firm support for my decisions while I was the demonstrator for the biotransformation practical. Throughout the three years of research, I also collaborated with Professor Helen Hailes and David Steadman from the Chemistry Department. Professor Hailes has been very dedicated and assertive which assured the direction and the progress of the project. David Steadman provided me not only the synthetic materials but also his time for discussions. This significantly improved my understanding of the chemistry side of the project which greatly facilitated my research. In sum, it was a great honour to work with such a fantastic research team.

My PhD would have never started smoothly without Leonardo Rios and Phattaraporn Morris who spent their time training me to safely use and maintain all the laboratory facilities. I would also like to thank Dhushy Stanislaus, Facilities Manager, who has put tremendous care and attention into maintaining laboratory safety, ensuring a good working environment and quickly resolving laboratory issues. I really appreciated her support for my role as lab manager during my final year. This was a once in a lifetime experience that I would never be able to find in day-to-day research activities.

The opportunity to study abroad was financially sponsored by the Royal Thai Government and I feel truly thankful to all Thai tax payers. I would also like to thank the Biochemical Engineering Department for additional funding for my international conference attendances. I would like to mention that I would never be able to apply for this degree without the support and recommendations of both Professor Gary Lye and Dr. Paul Dalby.

Throughout my life, my family has consistently supported me and I would never have achieved what I have without them. I feel especially grateful to my parents, Pichai and Sumalee Payongsri. I would also like to thank my brothers and sister, Nut, Srisakul and Titipol Payongsri, for their advice and moral support which has always lifted my spirits. Lastly, I would like to thank the Osters who made me feel that the UK is my second home.

Abstract

Transketolase catalyses asymmetric carbon-carbon bond formation and produces 1,3-dihydroxyketones, a functionality that is found in a vast number of natural and synthetic compounds. The wild-type transketolase enzymes from several species can accept a wide range of aldehydes, but industrially exploitable levels of activity tend to be limited to natural substrates and small aliphatic aldehydes. Several single mutants of the transketolase from *E. coli* were found to have enhanced activity towards non-phosphorylated substrates, non-hydroxylated aldehydes, and cyclic aldehydes. However, aromatic aldehydes still suffer from poor activities and yields while the key bottlenecks have not been identified. In order to improve transketolase activity towards aromatic aldehydes, small transketolase libraries were created with the aid of available crystal structures, and computational modelling of substrate binding. The new mutants were then assessed alongside transaminase for the ability to synthesise novel aromatic amino alcohols, which would provide building blocks for chloramphenicol and its derivatives.

The strategy for creating new libraries from combining single mutations can have significant impact not only on the activity but also the stability of the enzyme due to the synergy between residues. Random recombination of previously identified single mutants, which had higher activities on either glycolaldehyde or propionaldehyde, produced new double and triple mutants which all had lower activities for both substrates, and poor stabilities as illustrated by formation of inclusion bodies. On the other hand, the combination of D469 and R520 variants, two sites identified within a co-evolved network, created mutants that maintained the activity towards propionaldehyde, while significantly reducing the inclusion body formation. This implied that they were more stable and could be suitable templates for further rounds of mutation.

In the second stage, the factors influencing the bioconversion of aromatic aldehydes by transketolase were identified through crystal structure analysis. 3-formylbenzoic acid (3-FBA) and 4-formylbenzoic acid (4-FBA) were chosen as substrates that might re-establish hydrogen bonds between the carboxylate group and the so-called phosphate-binding residues within the active site. Both substrates were tested with wild type and a set of mutants that were previously reported to accept aromatic aldehydes. This showed that the steric hindrance from D469 impeded the aromatic substrate to access to the active site, and that D469T was the most active mutant for both new substrates. Site-directed mutagenesis at the phosphate-binding residues confirmed that R358, H461, and R520 interacted with carboxylate groups of both substrates and also suggested that the

affinity between the enzyme and the aromatic aldehyde, as well as their proximal orientations, was the key factor governing the reaction rate. This was also supported by the computational modelling of 3-FBA and 4-FBA binding in energy-minimised D469Tand D469T/R520Qstructures.

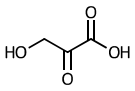
Site-saturation mutagenesis at S385 and R358 was performed to further improve the activity of transketolase for 3-formylbenzoic acid, 4-formylbenzoic acid, and also 3-hydroxybenzaldehyde (3-HBA). Due to the stability, D469T/R520Q was chosen as a template. S385E/D469T/R520Q was found with improved activity on 3-FBA and relieved substrate inhibition. Both S385Y/D469T/R520Q and S385T/D469T/R520Q were found to have higher activity and yields with both 4-FBA and 3-HBA. The enhancements in all the mutants were due to significant improvements in k_{cat}. Very large K_M values were observed for 3-HBA which confirmed that enzyme-substrate affinity was the major factor limiting the bioconversion rate. The synergic and kinetic study of S385Y/D469T and S385T/D469T illustrated that the triple mutants followed a different adaptive walk to reach the same optima.

In order to directly use 3-(1,3-Dihydroxy-2-oxopropyl) benzoic acid (3-DOPBA) and 4-(1,3-Dihydroxy-2-oxopropyl) benzoic acid (4-DOPBA) products from the transketolase reaction, as substrates in the transaminase reaction, with minimal competition from the leftover aldehydes, their conversion yields would have to be optimised. This was achieved by supplementing the reactions with an additional 50% HPA, which increased the yield for 3-DOPBA and 4-DOPBAto 90% and 70%, respectively. Both optimised reactions were fed into the transaminase reactions to synthesise novels aromatic aminodiols. However, none of the available transaminases appeared to accept either of the compounds. The competitive reaction between 4-FBA and 4-DOPBA in an amination reaction suggested that the major cause of the inability of CV2025 transaminase to catalyse 4-DOPBA was the inability of 4-DOPBA to access into the active site.

Abbreviations

3-DOPBA	3-(1,3-Dihydroxy-2-oxopropyl)benzoic acid
3-FBA	3-Formylbenzoic acid
3-HBA	3-hydroxybenzaldehyde
3-NBA	3-Nitrobenzaldehyde
4-DOPBA	4-(1,3-Dihydroxy-2-oxopropyl)benzoic acid
4-FBA	4-Formylbenzoic acid
AMP	Ampicillin
APD	(2S,3S)-2-aminopentane-1,3-diol
BA	Benzaldehyde
BSA	Bovine Serum Albumin
DHEThDP	Dihydroxyethyl-Thiamine diphosphate
DHP	1,3 dihydroxypentane-2one
DPP	1,3-dihydroxy-1-phenylpropane-2-one
E-4P	Erythrulose 4-phosphate
ee	enantiomeric excess
F6P	fructose 6-phosphate
GA	Glycolaldehyde
HPA	Hydroxypyruvic acid
HPLC	High-performance liquid chromatography
LB	Lysogenic broth
MBA	Methylbenzylamine
MW	Molecular Weight
PA	propionaldehyde
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PLP	pyridoxal 5-phosphate
PP domain	Phosphate binding domain
Pyr domain	pyrimidine binding domain
SDS	sodium dodecylsulphate
TAm	Transaminase
ThDP	Thiamine Diphosphate
TK	Transketolase
X5P	xylulose 5-phosphate

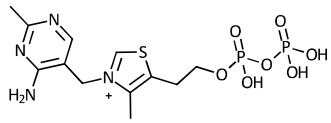
List of chemical structures



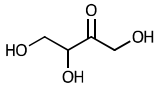
Hydroxypyruvic acid

Glycolaldehyde

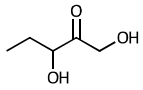
Propionaldehyde



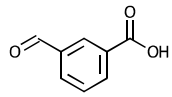
Thiamine diphosphate



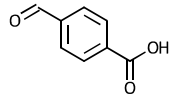
Erythrulose



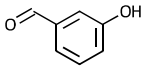
1,3 dihydroxypentane-2one



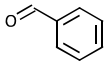
3-Formylbenzoic acid



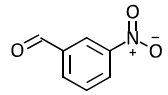
4-Formylbenzoic acid



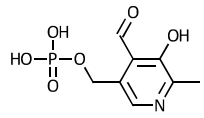
3-Hydroxybenzaldehyde



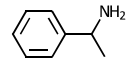
Benzaldehyde



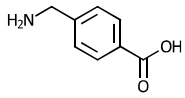
3-Nitrobenzaldehyde



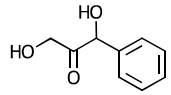
Pyridoxal 5-phosphate



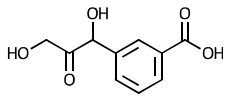
(S)-(−)-α-Methylbenzylamine



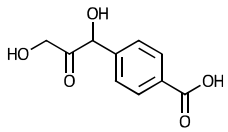
4-(Aminomethyl)benzoic acid



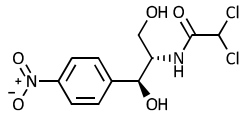
1,3-dihydroxy-1-phenylpropane-2-one



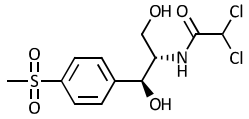
3-(1,3-Dihydroxy-2-oxopropyl)benzoic acid



4-(1,3-Dihydroxy-2-oxopropyl)benzoic acid



Chloramphenicol



Thiamphenicol

Table of Contents

List of Figure

List of Tables

List of equation

List of reaction scheme

1. Chapter 1: Literature review

1.1. Introduction

1.1.1. Biocatalysis

Enzymes have gained a huge interest in chemosynthesis due to their ability to catalyse reactions at low temperature with up to 10^9 folds improvement in the reaction rates (Koeller and Wong, 2001) together with high stereo- and regio-selectivity (Dalby, 2011; Koeller and Wong, 2001; Turner, 2009). These specificities are vital for the synthesis of pharmaceutical intermediates. Since they are able to catalyse their reaction with such high selectivity, the blocking and deblocking of functional groups can be avoided, which leads to fewer reaction steps (Ishige et al., 2005). The conditions required for the enzymes to function are often mild and involve less hazardous chemicals making the process more environmental friendly (Ishige et al., 2005) and minimise side reactions arisen from extreme pH and temperatures (Patel, 2008). If the harsh process conditions are unavoidable, the existence of extremophilic organisms have opened the possibilities of using biocatalysts in extreme conditions (Robertson and Steer, 2004).

Although biocatalysis is an attractive option, naturally exist variants may not have all the characteristics required for a particular industrial processes such as their stabilities in the required conditions, substrate spectrum and reactivity with non-natural substrates, and selectivity. Genomic approaches have been employed to find an enzyme with some desirable properties (Robertson and Steer, 2004) and protein engineering is often used to further improve other properties (Dalby, 2011).

The main focus of this research is to engineer a transketolase enzyme to accept aromatic aldehydes. This could provide an effective tool for the synthesis of chloramphenicol analogues. Therefore, this chapter focuses on the variety of protein engineering approaches available, and the background of previous research on transketolase. In the first half of this chapter, different methodologies of protein engineering were described together with the advantages and drawbacks of each approach. The second half of this chapter focuses on the mechanism of transketolase enzyme, the functions of certain residues within the active site, and the previous engineered transketolase variants and their substrate spectrum. The information from both

sections is vital for designing an approach fur further engineer transketolase. In the last section, the applications of transketolase in organic synthesis were reviewed.

1.1.2. Protein engineering: applications and strategies

Methods in protein engineering can be classified into two strategies; rational design, and directed evolution (Behrens et al., 2011; Lehmann and Wyss, 2001).

1.1.2.1. Rational design

Rational design requires sequence information as well as high quality structural information of the target enzyme and its closely related enzymes in order to select which residues should be replaced (Bornscheuer and Pohl, 2001; Cedrone et al., 2000). In recent years, this approach has also begun to employ computational modelling in order to justify the target, and predict the effect of the mutation prior introducing the mutation (Behrens et al., 2011). The ultimate *holy grail* of rational design is *de novo* enzyme design (Bolon et al., 2002; Otten et al., 2010). Since several enzymes can adopt the same fold while catalysing different overall reactions, different sets of catalytic residues within an active site could also generate several enzymes with diverse mechanisms (Glasner et al., 2006). Such a principle could be exploited in *de novo* enzyme design. Such a *de novo* design was achieved by the aid of computational modelling where force-field was applied to protein sequences in order to identify which sequence and conformation is the most stable for a particular backbone structure. This sequence is then used as a scaffold protein (Poole and Ranganathan, 2006; Street and Mayo, 1999). The positions and type of catalytic residues are then calculated and placed to generate active site (Bolon et al., 2002; Dalby, 2007). The use of computational modelling allows a vast sequence space search and narrow the library size by eliminating the sequences that are unlikely to fold (Chica et al., 2005). An alternative strategy to construct a novel enzyme is to design the positions and orientations of the catalytic residues while the scaffold proteins that could support these requirements are searched. Once the scaffold is found, the catalytic residues are then grafted into the scaffold (Röthlisberger et al., 2008). An enzyme catalysing a Kemp elimination, a reaction mechanism that does not occur in natural enzymes, has been successfully constructed by this method with a k_{cat} of 0.29 s^{-1}. Directed evolution has further improved this k_{cat} to 1.37 s^{-1} (Röthlisberger et al., 2008).

Amino acid substitutions are normally achieved by site directed mutagenesis where the mutations are incorporated into the primers which are then used in polymerase chain

reaction to synthesise a new DNA with the replaced codon. The amino acid can be replaced, deleted, or inserted depending on the primer design. Rational design has been shown to be successful in several cases. The thermostability of a protein can be improved by introducing more proline residues, salt bridge, and side chain-side chain hydrogen bonds but these are not universal rules (Lehmann and Wyss, 2001). Oxidation tolerance could be improved by the removal of cysteine and methionine (Bornscheuer and Pohl, 2001). The rational design approach can also improve the enantiomeric excess of an enzyme (Otten et al., 2010). In addition to improving the properties of an enzyme, rational design is usually employed when the relationship between amino acid residues and their function need to be determined.

The library size required for structural-based rational design tends to be smaller than that of directed evolution discussed in the next section. Consequently, less screening effort is required in theory (Kazlauskas and Bornscheuer, 2009). However, the limitation of rational design methods is the requirement of extensive information on the structure and sequences of enzymes. The unavailability of the structure could be overcome by computational modelling (Behrens et al., 2011). However, the dynamic and flexible nature of the protein structure still brings difficulties to rational design. Conformational changes upon mutation are rather difficult to predict and a small degree of differences can dramatically influence enzyme function. Therefore, it is extremely difficult to predict the outcome (Romero and Arnold, 2009).

1.1.2.2. Directed evolution

In contrast to rational design, directed evolution approaches do not require structural information. The process of directed evolution has been compared with natural evolution where the fitter species under a certain conditions is the most successful at passing on their genes to their offspring. The evolution of proteins could be viewed in a similar way where their function is expressed as their fitness. The mutation is artificially introduced into the gene and the natural selection occurs in the laboratory where the mutants are screened and selected. It could be viewed that different mutants have different degrees of fitness and mapping these sequences with their fitness has been described as fitness landscape Figure 1.1. In this map, each single sequence differing by one amino acid is aligned next to each other forming sequence space. Each of them has different fitness or its ability to function (Smith, 1970). This sequence space also illustrates evolutionary paths that a functional protein can take by acquiring one mutation at a time

with improving fitness. Under the screening or selection pressure, only fitter mutants are picked for further mutation which can be viewed as walking up the functional hill. This has been described as adaptive walk (Romero and Arnold, 2009). This sequence space is vast due to the fact that each position along the sequence can be assigned by 20 amino acids. For a small peptide of 100-amino acid long, there are already 100^{20} possible sequences. However, the majority of these sequences are not functional protein whereas the functional proteins tend to lie next to each other in the sequence space making hills in the protein sequence space (Smith, 1970).

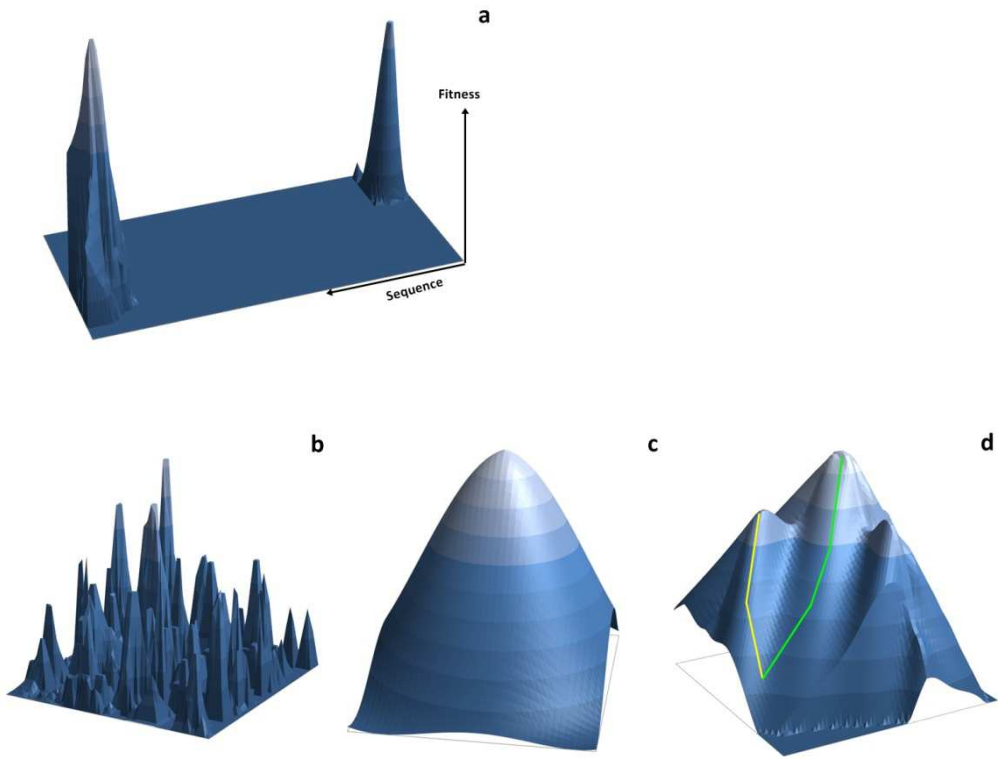


Figure 1.1 The Fitness landscape of a protein.

This figure illustrates the relationship between the sequence space and the activity of the enzyme which is represented by the intensity of the colour as well as the height of the structure. Higher and paler point illustrates higher activity. a) It is usually believed that among the vast sequence, only a small set of sequences is functional proteins. In reality, non-functional sequences are often found adjacent to the functional ones. b) Functional and non-functional sequences are often clustered together due to the epistatic interactions between amino acids which produce "Badlands" landscape. These interactions make it impossible to jump from one local optimum to another when the stepwise mutations are introduced. c) A Fujiyama landscape is less rough and the optimum can be easily reached than b). d) Multiple routes (yellow and green) can be found to reach the optimal point while certain path (yellow) may lead to local optimum.

The process of directed evolution starts with selecting a template protein with the some desired functions. This parental protein is then mutated to create a library of mutants

whose functions are assessed by screening or selection in order to identify any variants that have the desired properties. These mutants become the new templates and are subject to further rounds of mutations and screenings until the desired function is achieved.

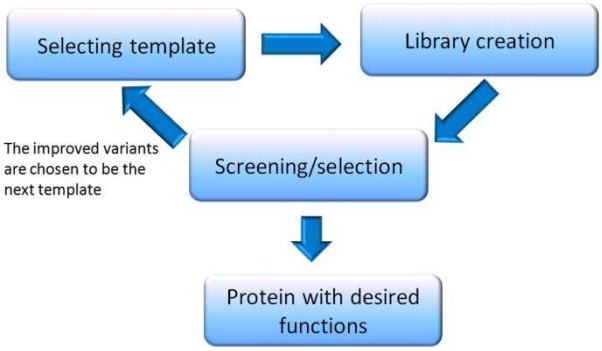


Figure 1.2 The process of directed evolution.

Starting with a good choice of parental protein could increase the chances of success with directed evolution. A good parental protein should have the desired function to a certain degree. However, mutations could destabilise the protein, so a stable parental protein is also preferable and can make protein more evolvable (Bloom et al., 2006; Tokuriki and Tawfik, 2009).

Once the parent is chosen, a library can be created. The strategies that are normally used to introduce mutations are error prone PCR (Cadwell and Joyce, 1992), recombinant techniques such as DNA shuffling (Stemmer, 1994a, 1994b) and staggered extension process (StEP) (Zhao et al., 1998). These recombinant techniques, however, require high sequence homology (70%) and the crossovers only occur in the regions with the highest sequence identity (Sieber et al., 2001). In addition, residues around the active site and substrate binding region tend to be highly conserved, so the chance that they are replaced by recombination is low (Paramesvaran et al., 2009). As a result, the mutation rate is still low and new functional proteins are less likely to be created while the library is also rather biased (Otten et al., 2010). In order to overcome these problems in the early recombinant techniques, several recombinant strategies that do not require sequence homology were developed. These include sequence homology–independent protein recombination (SHIPREC) (Sieber et al., 2001), incremental truncation for the creation of hybrid enzymes (ITCHY) (Ostermeier et al., 1999), a combination of ITCHY and DNA shuffling technique called SCRATCHY (Lutz et al., 2001) and several others reviewed by (Otten and Quax, 2005).

Screening or selection processes are methods for assessing the function of individual mutants within the library in order to identify better variants. The rule “you get what you screen for” (You and Arnold, 1996) must always be considered when designing the screening assay in order that the screening result can represent the function of interest as much as possible and avoid screening for something else. A large library requires the ability to individually culture a large number of variants as well as a high throughput screening method to analyse these variants. Screening technologies available include agar plates, microtitre plates, cell in droplet, cell as microreactor, cell surface display, and in vitro compartmentalisation (Leemhuis et al., 2009; Yang and Withers, 2009). The first two technologies are capable to work with the library size not more than 10^5 variants per day whereas the last four techniques are normally coupled to fluorescence-activated cell sorting system (FACS) and can handle between 10^9 - 10^{10} variants per day (Leemhuis et al., 2009). Most of the high throughput assays are coupled with absorption detection methods, either plate reader or FACS. The use of chromophoric or fluorophoric substrate or product could ease the process. However, the use of surrogate substrates or the indirect measurement of product such as derivatives could give inaccurate screening results. This is because the screened enzymes may evolve to work with the surrogate substrate instead of the actual substrate (Romero and Arnold, 2009). Despite the risk of inaccuracy, surrogate substrates still have other applications. If the protein is evolved to accept a new substrate that is structurally very different from the natural substrate, it could be impossible to screen the desired substrate directly. Instead, surrogate substrates or intermediates that their structures are gradually changed to the desired substrates could be used for stepwise screening (Arnold, 1998; Chen and Zhao, 2005; Savile et al., 2010).

Directed evolution has repeatedly shown to be a powerful tool for protein engineering without neither structure information nor sequence-function relationship. Apart from this purpose, it also shows several alternative adaptive pathways together with their probability of each pathway to occur in natural evolution (Romero and Arnold, 2009). For a functional multiple-mutation protein where one mutation can be introduced at a time, there are several routes to introduce each mutation. Any pathway that encounters deleterious mutations which are recovered by other mutations due to epistatic interaction is unlikely to be found in natural evolution. Therefore, directed evolution can suggest possible evolution pathway and mechanism. Epistatic interaction between residues or properties can also be studied (Romero and Arnold, 2009; Smith, 1970)

Substitution of one base in a codon does not always replace an amino acid due to the degeneracy of the genetic code. On average, one base pair replacement can result in 5.7 alternative amino acids (Kuchner and Arnold, 1997). Therefore, synonymous sequences (DNA variants giving identical protein sequence) can often be found within the library population, and the number of unique sequences is effectively lower. As a result, the number of samples to be screened has to be much larger in order to find all possible unique sequences (Drummond et al., 2005). In addition to degenerative code, one substitution tends not to improve the protein function. High mutation rate could, therefore, introduce multiple mutations. This comes with an expensive cost of even larger sequence space, hence, the library size (Drummond et al., 2005). If sufficient number of variants is screened, the chance of finding an improved protein is higher. All these complications make directed evolution labour intensive and require effective high throughput screening method. In addition, developing the screening method is not always easy and the properties or functions of interest may not be easily screened (Chica et al., 2005). The capacity of the high throughput screening still only cover a small fraction of the sequence space (Lutz, 2010).

1.1.2.3. Semi-rational design

As mentioned above, directed evolution tends to create a large library which requires intensive labour automated machinery while the accuracy of the screening can be doubtful (Romero and Arnold, 2009). Therefore, smaller and higher quality library construction can be preferable (Lutz, 2010).This strategy is often called semi-rational design or a smart library (Dalby, 2003; Lutz, 2010). Such strategies exploit the structure-function relationship, protein sequence, computational searches and predictive algorithms, to assist the prediction of the target sites in order to reduce the library size (Lutz, 2010). Site directed saturation mutagenesis is then introduced to those target sites. This could be used in combination with error-prone polymerase chain reaction to increase the mutation rate at other sites in addition to the target residue (Chica et al., 2005). Four main redesign strategies are often employed to find the targets and these are sequence-based, structure-based, computational-based redesign, and computational *de novo* enzyme design as reviewed by (Lutz, 2010). A combination of several strategies is also possible.

Even though the library can be narrowed down due to the reduction in the target sites, multi-site saturation mutagenesis can still give very large number of possible combinations. Degenerative codon is still an issue but this can be minimised by using NDT

codon which codes for 12 amino acids (Reetz and Carballeira, 2007) instead of NNH that codes for 20 amino acids. The size of the library could be minimised by employing *in silico* screening of all mutation combinations (Hayes et al., 2002). This screening ranks the mutants according to their global minimum energy conformation which is associated with folding and protein stability. The mutants predicted to be stable are then created in the laboratory and their functions are experimentally assessed. Although mutations at the residues close to the catalytic centre have more potential to improve enzyme activity, choosing these residues is not always easy. One approach to identify potential targets is to calculate sequence entropy which represents the degree of sequence conservation. Mutation at the site with high entropy is more likely to improve the function of an enzyme (You and Arnold, 1996). Besides sequence entropy analysis, structure, functional and evolutionary databases could all be used to predict the target sites as shown in HotSpot Wizard (Pavelka et al., 2009). The hot spots are highly variable functional residues which their replacements have a tendency to change the enzyme activity (Pavelka et al., 2009). This wizard also annotates highly conserved catalytic residues, ranks the mutability of each residue on both sequence and structure. However, the improvement of enzymatic activity and selectivity after mutation may not be significant compared with mutations at the moderate to highly conserved residues which have been shown to be successful in several cases (Morley and Kazlauskas, 2005; Paramesvaran et al., 2009; Toscano et al., 2007).

Semi-rational design can provide several advantages over the traditional directed evolution. These include but not limited to the non-requirement of high throughput screening and more accurate method can be used for screening (Lutz, 2010).

1.1.3. Evaluation of the kinetic parameters for organic synthesis

Protein engineering can improve several properties of an enzyme to meet the performance required for a particular process condition. These properties are the stability, k_{cat}, Michaelis-Menten constant (K_M), and substrate and product inhibition constants (Fox and Clay, 2009; Koeller and Wong, 2001).

Traditionally, the catalytic efficiency of an enzyme is expressed as k_{cat}/K_M and the higher the value is, the higher the efficiency. In addition, this value is also used to compare the specificity of an enzyme towards different substrates. However, during protein engineering, two enzymes may have been identified to have very similar k_{cat}/K_M values but their individual k_{cat} and K_M values may be significantly different. This becomes a question

which enzyme is more suitable for the process. Several works have shown that the term k_{cat}/K_M is a rather meaningless and shall not be used to compare enzyme efficiency (Ceccarelli et al., 2008; Eisenthal et al., 2007; Fox and Clay, 2009). This term, however, excludes several factors that are important for space-time yield consideration (Fox and Clay, 2009). In industrial process where enzymes are used, substrate concentrations usually exceed what are found in physiological conditions which can cause enzyme inhibition. Considering Michaelis-Menten equation below, it could be seen that at extremely high substrate concentration where $K_M << [S]$, the reaction is operating solely depending on the k_{cat} and K_M is much less important.

$$V = \frac{k_{cat} E_t [S]}{K_M + [S]}$$

Equation 1.1 Michaelis-Menten equation. E_t is the total enzyme concentration, V is velocity, [S] is substrate concentration

The enzyme with the highest k_{cat} will perform faster, hence, reaching the highest conversion and productivity. This also means the highest substrate concentration possible should be used during the screening and the time required to reach the highest conversion of each individual enzyme should be determined (Fox and Clay, 2009). However, the k_{cat}/K_M values towards different substrates can still be used to compare substrate preference or specificity only when the concentrations of both substrates do not exceed their K_M(Ceccarelli et al., 2008). This further limits the application of this term in industrial processes, especially in one-pot synthesis where all the highly concentrated substrates are added simultaneously and an enzyme can react with several substrates.

Since the k_{cat}/K_M value is not an appropriate parameter to justify which enzyme should be used in an process, Fox and Clay (2009) suggested to compare different enzymes by their average velocity which integrated inhibition constants and other kinetic parameters into rate equation (Fox and Clay, 2009). This equation is applicable for batch and fed-batch reaction systems.

Once an enzyme with enhanced activity is created, the next consideration is how good is good enough. This depends on the turnover required by the process. A k_{cat} of 2 s^{-1} can generate 1 mmol of product per day (Koeller and Wong, 2001). Although the above enzyme efficiency may discard the K_M value from its consideration, the low K_M can cause substrate inhibition.

1.1.4. Transketolase

Transketolase (E.C. 2.2.1.1) belongs to thiamine diphosphate (ThDP) dependent enzyme family. It catalyses an asymmetric reversible transfer of a dihydroxyethyl group from a ketose phosphate to an aldose phosphate in non-oxidative phase of pentose phosphate pathway and Calvin cycle (Kochetov and Sevostyanova, 2005; Schenk et al., 1998). Transketolases have been shown to have high stereospecificity and selectivity. They prefer hydroxylated aldehydes with a 2*R* configuration and the transfer of the dihydroxyethyl group to a 2*R* hydroxylated aldehyde produces a dihydroxy ketone with a 3*S*,4*R* or *D-threo* configuration (Hecquet et al., 1994; Hobbs et al., 1993; Turner, 2000)(see Scheme 1.1). The use of hydroxypyruvate (HPA) as a ketol donor leads to the production of carbon dioxide which pulls the reaction to completion (Scheme 1.2).

Scheme 1.1 General reaction catalysed by transketolase.

Scheme 1.2 The reaction of transketolase when HPA is used as a substrate.

Molecular evolutionary analysis revealed that several other enzymes belong to this family and they have been classified into six groups according to the domain arrangements and additional domain recruitments within an enzyme (Costelloe et al., 2008). These six groups are transketolase-like, pyruvate ferredoxin reductase, pyruvate decarboxylase-like, 2-oxoisovalerate dehydrogenase-like, sulfopyruvate decarboxylase, and phosphopyruvate decarboxylase (Costelloe et al., 2008). All the members in these groups have two highly conserved domains namely pyrophosphate binding domain (PP) and pyrimidine binding domain (Pyr). In most of these members, both domains are found on the same subunits except for 2-oxoisovalerate dehydrogenase like enzymes (Costelloe et al., 2008). The multiple sequence alignment of the PP and Pyr domains of the enzymes within this family revealed several highly conserved residues. Some residues are found to be conserved only within a few groups but the residues that are responsible for metal binding, ThDP binding, and ThDP activation are conserved in all six groups (Costelloe et al., 2008). In addition to these highly conserved residues, multiple sequence alignment of ThDP dependent enzymes

also revealed that there is a common ThDP binding motif which consists of $GDGX_{26}NN$ or $GDGX_{26}NCN$ (Hawkins et al., 1989). This sequence adopts αβα fold and involves in the binding of divalent cations (Lindqvist and Schneider, 1993; Nemeria et al., 2009) which in turn bind to the diphosphate group of the ThDP molecule.

1.1.4.1. Crystal structures of transketolases

In most species, transketolases exist as a homodimer while transketolase from spinach and *Candida boidinii* exist as monomer and tetramer, respectively (Schenk et al., 1997). In yeast and *E. coli*, the two ThDP molecules and two divalent cations per transketolase molecule are required for their function (Figure 1.3). The crystal structures of transketolase from both species revealed that ThDP molecules bind at the active site interface (Asztalos et al., 2007; Nikkola et al., 1994). The binding of ThDP to transketolase molecule brings several changes to both the ThDP molecule itself and transeketolase in terms of structure and function.

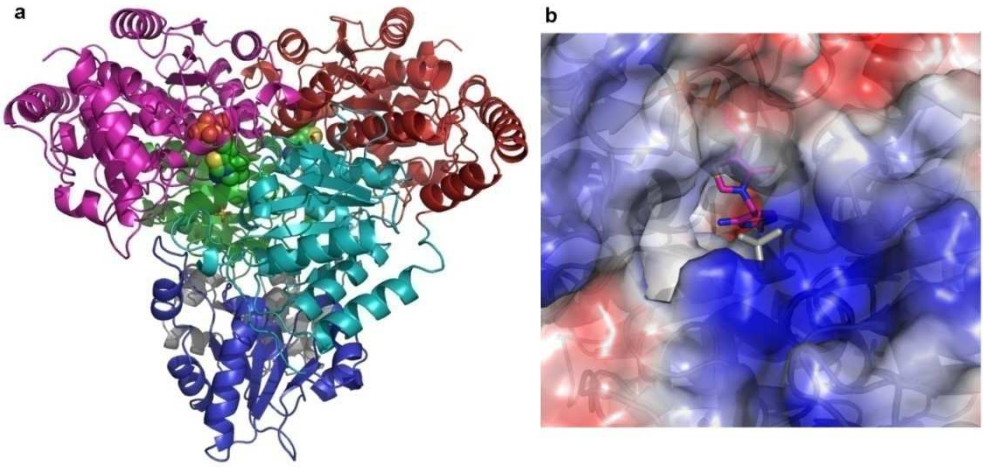


Figure 1.3 Transketolase structure from *E. coli* (PDB ID: 1QGD).

a) The ribbon structure. The PP, Pyr, and C term domains of the subunit one are in magenta, green, and grey respectively. The PP, Pyr, and C term domains of the subunit two are in red, cyan, and blue respectively. Both ThDP molecules are shownas spheres. b) The surface above the active site. Only the C2 of ThDP is accessible to solvent.

1.1.4.2. Mechanism of transketolase

1.1.4.2.1. The role of ThDP in transketolase reaction

ThDP is a derivative of vitamin B_1. The structure consists of one pyrimidine ring linked to a thiazolium ring, and a diphosphate group. The C2 atom is the centre of all ThDP dependent enzymes. In transketolase, this is the only atom that can be accessed by solvent (see also Figure 1.3B) (Nikkola et al., 1994; Sundström et al., 1992). The conformation of ThDP molecule in free solution differs from the enzyme-bound where the 4'-NH_2 is placed next to the C2 atom. The latter conformation is called the "V conformation". Both conformations are shown in Figure 1.4 below.

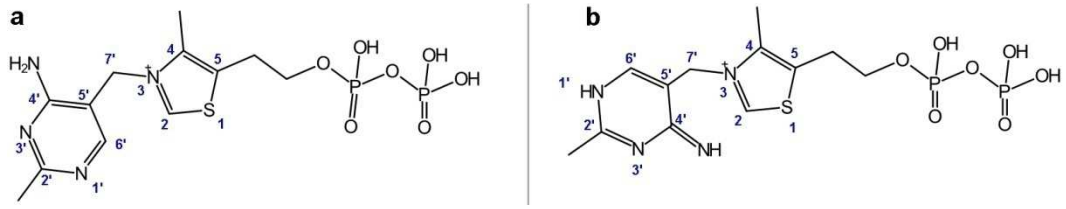


Figure 1.4 Two conformations of ThDP.

a) inactive conformation. b) the V conformation.

ThDP is directly involved in the catalytic reaction of all ThDP dependent enzymes without being released from the active site (Kluger and Tittmann, 2008). Free thiamine and thiazolium derivatives can catalyse similar reactions observed in the ThDP dependent enzyme but with several orders of magnitude slower (Kluger and Tittmann, 2008; Mizuhara and Handler, 1954; Schenk et al., 1998). The mechanism of ThDP has long been studied but the early researches proposed that the amine group on the pyrimidine formed a Schiff base with the carbonyl group(Langenbeck, 1932; Wiesner and Valenta, 1956). However, it was illustrated that this amine cannot be deprotonated (Breslow, 1958; Fry et al., 1957).Therefore, the formation of a Schiff base by the amine on the pyrimidine and mechanism involved are rather impossible. Breslow (1957) used NMR spectroscopy to monitor the deuterium exchange rate of ThDP and revealed for the first time that the C2 of the thiazolium ring can be deprotonated. He then proposed the catalytic mechanism of ThDP where the deprototanted C2 forms a carbanion and acts as the nucleophile that attacks the carbonyl group(Breslow, 1958). The enzymes within this family employ this mechanism for all their substrates (Kern et al., 1997). The formation of this carbanion has been investigated in terms of deprotonation rate by 1H NMR spectroscopy and the equilibrium position of the carbanion by the p*K*a value of this C2 atom. The H/D exchange

at the C2 position in solvent deuterium has a half-life of 20 minutes (Breslow, 1957)which is too small to catalyse any reaction (Kern et al., 1997). The p*K*a of this CH atom is between 17-20 (Kemp and O'Brien, 1970)which makes a very small fraction of ThDP exists in the carbanion form (Hübner et al., 1998). These evidences suggest these enzymes play a crucial role in accelerating the reaction but they do not catalyse the reaction themselves (Kluger and Tittmann, 2008).

The role of the enzyme in activating ThDP was revealed when Kern and colleagues investigated the exchange rate at the C2 atom with pyruvate decarboxylase and transketolase (Kern et al., 1997). These enzymes can increase the exchange rate up to three orders of magnitude higher than the free ThDP but most of the C2 atoms are still protonated. Pyruvate decarboxylase requires allosteric or substrate activations to further increase the exchange rate to meet its observed k_{cat}(Kern et al., 1997).

The crystal structures several ThDP dependent enzymes show the three essential roles of these enzymes in ThDP activation; the V conformation, the position of 4'-NH_2, and the glutamate side chain next to the N'1 atom (Schellenberger, 1998). Their crystal structures show that the ThDP molecule is in the V conformation which brings the 4'-NH_2 group in close proximity to the C2 atom (Figure 1.4). One glutamate that forms a hydrogen bond with the N'1 atom is highly conserved in all the ThDP dependent enzymes. Mutation at this residue in both pyruvate decarboxylase and transketolase from yeast results in severely impaired k_{cat}. This implies that this glutamate residue is involved in the activation of ThDP through the proton abstraction at the N1' atom which transfers the negative charge to the 4'-N (Hübner et al., 1998; Kern et al., 1997; Schellenberger, 1998). 4'-NH_2 was identified to act as the proton acceptor of the C2 atom through the use of 4'-desamino-ThDP and site directed mutagenesis at the histidine residue close to the C2. Replacing the histidine residue does not change the deprotonation rate of the C2 atom but removing the 4'-NH_2 significantly impairs the k_{cat} of both enzymes (Kern et al., 1997). In addition, the deprotonation rate of the C2 atom is at least 5 orders of magnitude lower (Hübner et al., 1998). This concludes that the 4'-NH_2 promotes the C2 deprotonation.

1.1.4.2.2. Reaction mechanism: active site residues and ThDP

In addition to activating ThDP, certain residues in the active site of transketolase were identified to act as acid/base catalysts which transfer protons between substrates as illustrated in the reaction mechanism in Figure 1.5 (Wikner et al., 1997).

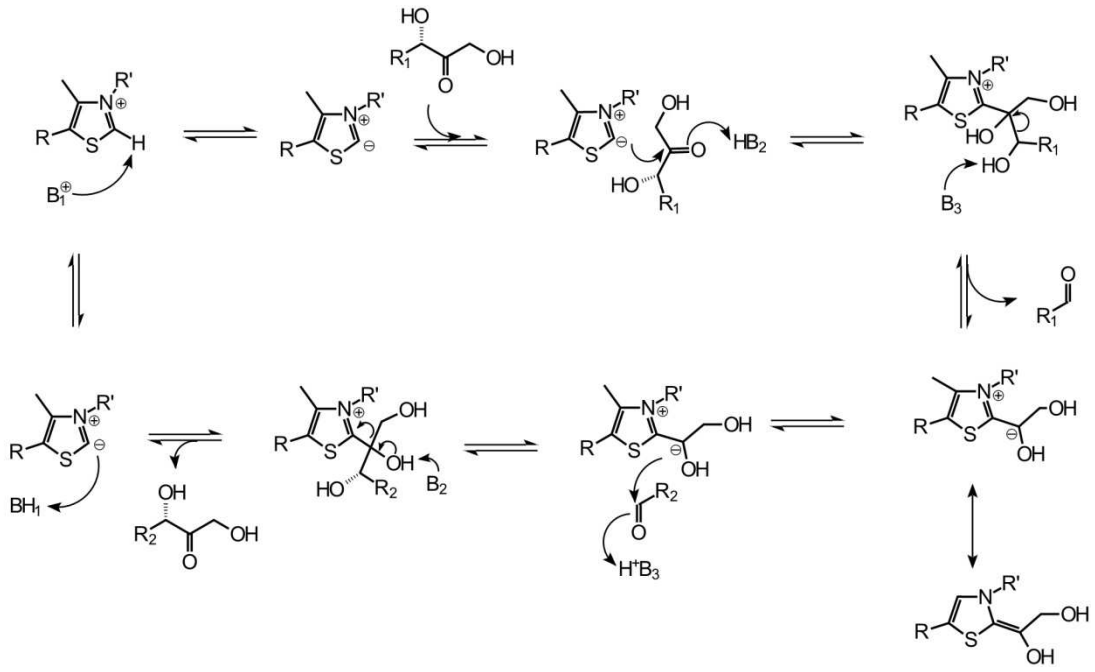


Figure 1.5 The Catalytic mechanism of transketolase.

R is the pyrimidine ring and R' is the rest of the ThDP molecule. B1, B2, and B3 are acid/base involved in the catalysis.

The reaction mechanism of transketolase is divided into two stages (Nilsson et al., 1997; Wikner et al., 1997) and follows the ping-pong mechanism. Both stages require proton transfers between ThDP, the active site residues, and the substrates. In the first stage, proton abstraction at the C2 atom by B_1 must occur to form a carbanion which then acts as a nucleophile that attacks the carbonyl group of the donor substrate. Delocalisation of electrons in the thiazolium ring of ThDP stabilises this carbanion and this mode of stabilisation is often described as electron sink. The carbonyl group is then protonated by B_2 to form a hydroxyl group. The proton abstraction at the C3 hydroxyl group of the intermediate by B_3 then leads to the carbon-carbon bond cleavage. At the end of this first stage, the carbanion of dihydroxyethyl-ThDP intermediate is formed and an aldehyde is released from the active site (Wikner et al., 1997). In the absence of an acceptor aldehyde, the dihydroxyethyl group can be released from ThDP and forms glycolaldehyde (Fiedler et al., 2001). The second stage is the ligation of the dihydroxyethyl group to an acceptor. The carbanion acts as a nucleophile attacking the carbonyl group of the aldehyde acceptor forming another covalent-linked intermediate. The later steps are the reverse process of the first stage and the release of the dihydroxy ketone from the active site. In both stages,

the oxygen on the carbonyl group of the donor and acceptor must be stabilised through either charge compensation or proton transfer (Nilsson et al., 1997; Wikner et al., 1997).

It can be seen from this mechanism that several acid/base catalysts are involved in the reaction. Enzyme crystallography, site-directed mutagenesis in yeast (Wikner et al., 1997) and *E. coli* transketolase (Asztalos et al., 2007), and the use of ThDP analogues were employed to identify these acid/base catalysts. In the active site of transketolase, there are four highly conserved histidine residues which can involve in the proton transfer; H26, H69, H261, on the PP domain, and H473 on the Pyr domain of *E. coli* transketolase as shown in Figure 1.6a.

As previously mentioned, 4'-NH_2 directly associates with the proton abstraction at the C2 atom when it forms 4'-imino-ThDP (Hübner et al., 1998; Kern et al., 1997). Therefore, this 4'-NH_2 was identified as the proton acceptor in this step. The carbonyl group of the donor molecule was suggested to be stabilised either by the H481 (H473 in *E. coli*) or 4'-imino group (Wikner et al., 1997). The crystal structure of yeast transketolase with erythrose 4-phosphate (PDB ID: 1NGS) suggested that the residues that abstract the proton of the donor C3 hydroxyl could be H30 and H263 due to the fact that these residues form hydrogen bonds with the carbonyl group of erythrulose 4-phosphate (Nilsson et al., 1997). Since erythrose is the product from the cleavage of fructose, the location of the carbonyl group of erythrose and the C3 hydroxyl group of fructose would be nearly identical (Nilsson et al., 1997). This proposal was further supported when the crystal structures of *E. coli* transketolase with xylulose-5-phosphate (PDB ID: 2R8O) and fructose-6-phosphate (2R8P) showed that the C3 hydroxyl groups of both substrates interact with H26 and H261 (Asztalos et al., 2007). This carbonyl group is formed upon the deprotonation of the C3 hydroxyl group of during the first stage of the reaction. In the double mutant H26A/H261A in *E. coli* transketolase, much fewer donor-ThDP adduct species also formed (Asztalos et al., 2007). These evidences lead to a suggestion that both histidine residues could also be involved in the proton transfer and be nominated as B3 in Figure 1.5(Nilsson et al., 1997).

1.1.4.2.3. Other residues involved in mechanism: substrate recognition and stereoselectivity

In addition to acid/base catalysts and ThDP activation, several highly conserved residues in transketolase active site and in vicinity serve other various functions including

substrate recognition, stereoselectivity control, transition state stabilisation, ThDP binding, divalent cation metal binding, and phosphate binding. Mutagenesis studies and the crystal structures of transketolases from both yeast and *E. coli* with natural substrates were used to interpret the functions of these residues. In addition, the crystal structure of yeast transketolase with enamine-ThDP intermediate was also available (1GPU). These studies revealed that some residues perform multiple functions at the same time through their complex hydrogen bond network with ThDP, substrates, and intermediate. Here, the position of each amino acid is displayed in Figure 1.6 a) and b) and the functions of each residue are described. All the residue positions refer to *E. coli* transketolase unless stated. The positions in yeast transketolase are in parenthesises when both species were mentioned together.

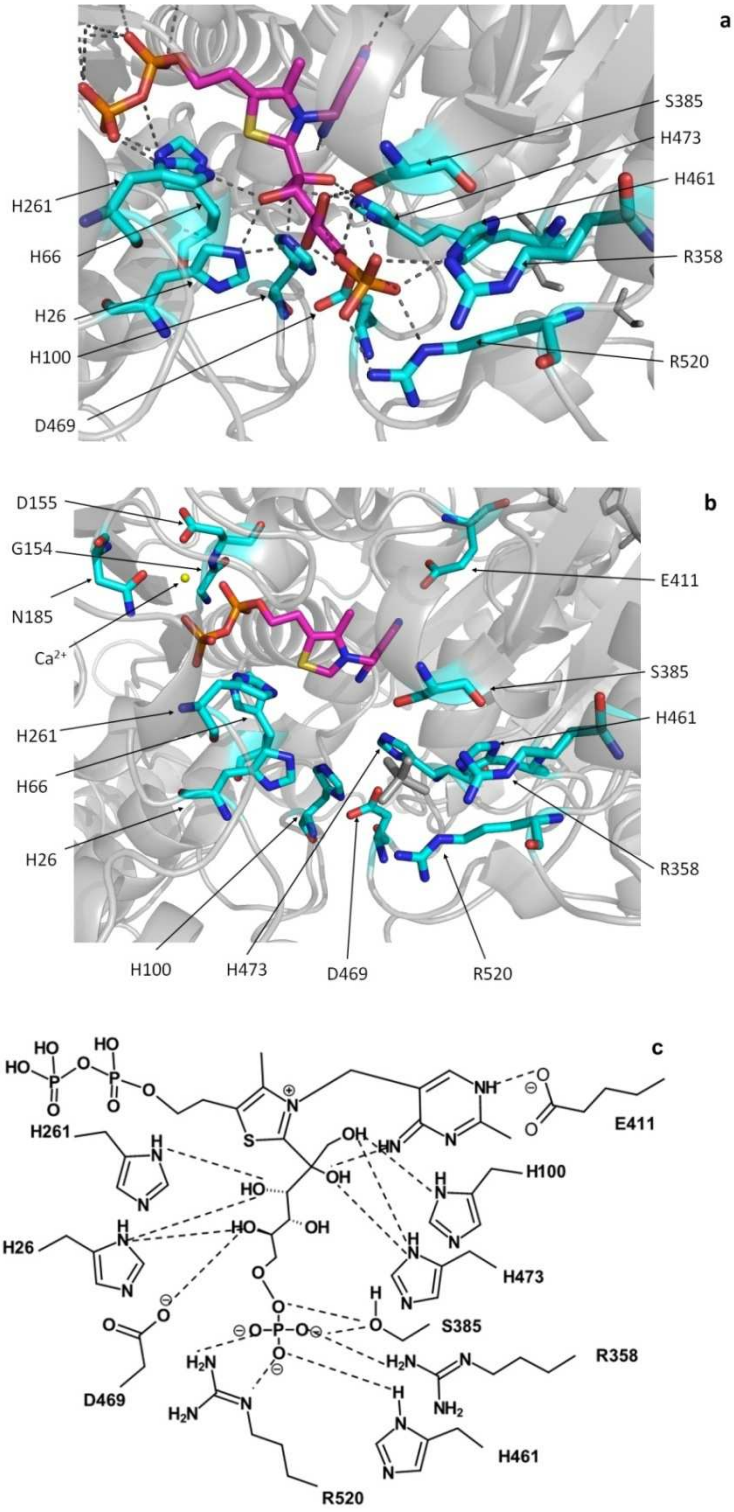


Figure 1.6 Residues within *E. coli* transketolase active site.

a) The hydrogen bond network between donor-ThDP adduct (PDB ID: 2R8P). b) The conserved residues important for ThDP binding, substrate recognition, and catalysis. c) A schematic draw representing the conserved residues involved in catalysis and substrate recognition. Hydrogen bond network was adapted from the *E. coli* TK crystal structure (PDI ID: 2R8P) and are shown as dotted lines. For a summary of the functions of each residue, see Table 1.1.

As previously mentioned, H30 and H263 in yeast transketolase were proposed to be part of the proton transfer in the catalytic step because the mutation at this residue in yeast transketolase significantly reduces the k_{cat} value (Wikner et al., 1997) and the formation of donor-ThDP adduct (Asztalos et al., 2007). In addition, mutation at H30 also increases the K_M value towards the donor xylulose-5-phosphate without noticeably altering the K_M towards the acceptor ribose-5-phosphate (Wikner et al., 1997). The different effects of the mutation upon the K_M values towards the donor and acceptor could arise from the fact that hydrogen bonds only form between H30 and the C3 hydroxyl group of the donor substrate but not with the acceptor. Therefore, removing this hydrogen bond can lead to a lower enzyme-donor substrate affinity but not the affinity towards the acceptor (Nilsson et al., 1997). However, a rather contradictory observation was found in *E. coli* transketolase where the wild type and the double mutation H26A/H261A have similar K_M values towards donor and acceptor substrates although the activity of the double mutant is less than 1% of the wild type (Asztalos et al., 2007). The crystal structure of *E. coli* TK with ribose-5-phosphate (PDB ID: 2R5N) also shows that the H26 forms hydrogen bond with the C3 hydroxyl group (Asztalos et al., 2007).

It was observed that the hydroxylated donor substrate must be in *D-threo* configuration at the C3 position (Hecquet et al., 1994; Morris et al., 1996) and the C3 hydroxyl group forms hydrogen bonds with H26. Therefore, the additional function of H26 was suggested to control the stereoselectivity of the enzyme. Later on, H26Y was identified to reverse the stereoselectivity from 3*S* to 3*R* with an enantiomeric excess greater than 88% when using propylaldehyde as an acceptor (Smith et al., 2008). This further supports the role of H26 in stereo-control and substrate recognition.

H66 is one of the conserved histidine residues within the active site but it locates close to the phosphate group of the ThDP molecule. The crystal structures of *E. coli* transketolase (PDB ID: 2R8P, 2R8O) show that one of the nitrogen atoms on the imidazole ring forms hydrogen bonds with the phosphate group of ThDP and the other nitrogen atom forms hydrogen bonds with a water molecule which in turn forms hydrogen bonds with the C1-hydroxyl group of the donor-ThDP adducts. The DHEThDP intermediate in yeast transketolase (PDB ID: 1GPU) also shows that the imidazole ring of H69 also forms hydrogen bond with the hydroxyl group of the DHEThDP. However, both crystal structures from *E. coli* (PDB ID: 2R5N) and yeast (PDB ID: 1NGS) do not show that this residue does not form hydrogen bonds with acceptor substrates. Mutation at H69 in yeast transketolase

resulted in 3 folds higher K_M towards ThDP and over 5 folds increase in the K_M towards the donor while the k_{cat} is also severely diminished (Wikner et al., 1997). The impact of the mutation on these catalytic parameters implies that these hydrogen bond networks could be important for the orientation of the donor-ThDP and the DHEThDP molecules (Wikner et al., 1997).

H100 could also be another residue that may involve in substrate recognition and intermediate stabilisation as identified in yeast transketolase (H103) (Wikner et al., 1995). When substituting this histidine in yeast transketolase with alanine, asparagines, or phenylalanine, it markedly reduces the enzyme activity with significant increase in K_M towards both ThDP and the donor molecule (Wikner et al., 1995) although phenylalanine is one of the natural variants at this position (Hibbert et al., 2007). The impact of the mutation on the K_M towards the donor molecule and the position of this residue suggested that this histidine also interacts with the substrate molecule (Wikner et al., 1995). This was confirmed later by the crystal structures of *E. coli* transketolase (PDB ID: 2R8O and 2R8P) that, indeed, H100 forms hydrogen bonds with the C1-hydroxyl group of the donor-ThDP adduct molecule and could stabilise this intermediate (Asztalos et al., 2007).

The enzyme-substrate interactions that have been mentioned so far are those between the hydroxyl groups of the substrates and the highly conserved histidine residues. The natural phosphorylated substrates also bind to the active site through phosphate binding residues which are R358, H461, and R520 (R359, H469 and R528 in yeast transketolase) (Asztalos et al., 2007; Nilsson et al., 1997). An additional residue in *E. coli* transketolase S385 also takes part in the binding of fructose-6-phosphate and xylulose-5-phosphate (Asztalos et al., 2007). Replacing one of the positively charged residues with alanine reduces the affinity between the enzyme and phosphorylated substrates (Nilsson et al., 1997). Phosphorylated substrates have the K_M values between 0.1 – 7 mM whereas the K_M values towards the non-phosphorylated sugars are up to three order of magnitude higher than those phosphorylated ones (Sprenger et al., 1995). Although hydroxypyruvate, an extensively used donor molecule, lacks the phosphate group, it may still bind to the active site via the carboxylic acid group. However, the study in yeast and *E. coli* transketolase illustrated that the K_M values of hydroxypyruvate are up to two orders of magnitude higher than the phosphorylated ketoses (Schenk et al., 1998; Sprenger et al., 1995). The non-phosphorylated substrate, glycolaldehyde, also has the K_M value of 35 mM (Hibbert et al., 2007) which is also an order of magnitude higher than the natural

phosphorylated substrates. This emphasised the strength of the substrate binding through the phosphate group and how it can have an impact on the substrate conversion rate. However, mutation at these residues have been shown to improve the activity towards glycolaldehyde (Hibbert et al., 2007).

As previously mention, transketolase is highly stereoselective and this is controlled by the hydrogen bond network formed between H26, H261, and the substrate. In addition to these two residues, D469 (D477 in yeast transketolase) was also found to interact with the C3 and C4 hydroxyl group of the donor and C2 hydroxyl group of an acceptor (Asztalos et al., 2007; Nilsson et al., 1998, 1997). The importance of the C2 hydroxyl group in acceptor molecules was demonstrated by removing the C2 hydroxyl group from the acceptor substrates or replacing this aspartate with alanine. In both cases, very low substrate conversion was observed together with a significant increase in k_{cat}/K_M(Nilsson et al., 1998). Transketolase from yeast and *E. coli* also poorly accept alternative diastereomers which have the opposite configuration at the C2 position (Morris et al., 1996; Nilsson et al., 1998). Although the interaction between the C2-hydroxyl group and D469 seems to be crucial, several D469 mutants have been shown to improve specific activity towards propionaldehyde (Hibbert et al., 2008) with improved and reversed stereoselectivity (Smith et al., 2008), longer chain aliphatic aldehydes with higher enantiomeric excess than wild type (Cázares et al., 2010), and even aromatic aldehydes (Galman et al., 2010; Payongsri et al., 2012).

During the catalysis of transketolase, the DHEThDP intermediate has to be stabilised and the stabilising functions were proposed to arise from H473. The crystal structure of yeast transketolase with DHEThDP (1GPU) revealed that H481 forms hydrogen bond with the α hydroxyl group of the DHEThDP in addition to other histidine residues (Schneider et al., 2002). This residue is also in close proximity to the 4'-NH_2 group, and it was proposed that H473 may accept a proton from 4'-NH_2 to form 4'-imino group to abstract the C2-proton (Singleton et al., 1996). However, amino acid substitution at this residue does not influence the deprotonation rate of the C2 atom on the thiazolium ring suggesting that it does not promote the ThDP activation (Hübner et al., 1998). In addition, H473 is only conserved in non-mammalian species (Wikner et al., 1997) and only some ThDP dependent enzymes (Costelloe et al., 2008). In human transketolase, this residue is replaced with glutamine and it has been shown a higher degree of tolerating amino acid substitution (Singleton et al., 1996) compared with H481 mutation in yeast transketolase

(Wikner et al., 1997). It should be noted that the human transketolase shows 24% sequence identity to yeast transketolase and the variation in the sequence may lead to the alteration in the function of the residue. This leads to a suggestion that H110 in human transketolase (H100 in *E. coli*) may serve the function of H473 instead (Singleton et al., 1996).

As illustrated in Figure 1.3, each ThDP molecule binds at the interface of the two subunits where the phosphate group binds to the PP domain of subunit one through the interaction with divalent metal ion and the pyrimidine ring binds to a hydrophobic pocket in the Pyr domain of the other subunit. The thiazolium ring forms a hydrophobic interaction with both subunits (Schneider and Lindqvist, 1998). Upon the ThDP binding, two disordered regions undergo structural rearrangement and become more ordered. These regions are then identified as cofactor loops (Sundström et al., 1992). In yeast transketolase, the two regions consist of N187-W198 and L383-G398 (Sundström et al., 1992) whereas in *E. coli* transketolase, these two regions are N185-W196 and L382-G392 (Martinez-Torres et al., 2007). The divalent metal ion forms electrostatic interaction with the side chain of D155 (D157), N185 (N187), and the main chain oxygen of I187 (I189) (Sundström et al., 1992). The loop closure is stabilised through several inter-subunit interactions and the second loop forms hydrophobic interaction with the first loop (Sundström et al., 1992). Besides the interaction between two loops, an additional aspartate residue also involves in the loop closing. The amino acid substitution at D382 in yeast transketolase leads to the second loop disorder which suggested that D382 stabilises the loop formation although this residue is not part of the cofactor loop (Meshalkina et al., 1997). In fact, this residue involves in subunit-subunit interaction by forming hydrogen bond with the first loop of the other subunit (Sundström et al., 1992). The loss of this interaction leads to a further loss of loop-loop interaction and second loop disorder. Besides that, substituting this aspartate for alanine or asparagine reduces the enzymatic activity and increases the K_M towards both phosphorylated donor and acceptor molecules (Meshalkina et al., 1997). In addition to providing the inter-subunit interaction, it was observed that ThDP binding and the loop closure influence the thermal stability of transketolase (Esakova et al., 2005).

The dimeric enzyme is not only stabilised through the interaction between the cofactor loops. The study in yeast transketolase illustrate that the two subunits are also held together by hydrogen bond network formed by highly conserved acid residues; E162 of the first subunit with G167 and E418 of the other subunit (Meshalkina et al., 1997).

These acid residues are, in fact, found to be conserved in all ThDP dependent enzyme family (Costelloe et al., 2008; Frank et al., 2004). Later, additional function of these acid residues was suggested to act as proton wire which alternates the protonation state of the ThDP in the adjacent active site in all ThDP dependent enzyme family (Frank et al., 2004).

Table 1.1 The summary of the functions of each residue.

Residue	Function	Residue	Function
H26	Acid/base catalysis (B_3), stereospecific	R358	Bind to Pi of substrate
H66	Substrate recognition, bind to phosphate group of ThDP	D381	ThDP loop closure
H100	Substrate recognition, intermediate stabilisation	S385	Bind to Pi of substrate
D155	Divalent metal binding	E411	Active ThDP by proton abstraction at the N1' atom
E160	H-bond network in proton wire	H461	Bind to Pi of substrate
E165	H-bond network in proton wire	D469	Stereospecific
N185	Divalent metal binding	H473	stabilise transition state
H261	Acid/base catalysis (B_3), stereospecific	R520	Phosphate binding

1.1.4.3. Application of transketolase in organic synthesis

1.1.4.3.1. Asymmetric carbon-carbon bond elongation: the advantages of transketolase

Asymmetric carbon bond elongation is a powerful tool in organic synthesis which enables the construction of a building backbone of natural and synthetic molecules (Breuer and Hauer, 2003; Fessner, 1998; Machajewski and Wong, 2000; Resch et al., 2011; Wohlgemuth, 2009). Some chemosynthetic methods are available for the synthesis of dihydroxyketones via carbon-elongation but most of them are multi-step and still suffer from poor atom efficient (Hailes et al., 2009) and the instability of the intermediate hydroxyl carbanion equivalent (Hailes et al., 2010; Smith et al., 2006). Only recently, one-step synthesis of racemic dihydroxyketones with greater atom efficient is available by using 3-(N-morpholino)propanesulfonic acid as a catalyst (Smith et al., 2006). The mechanism was further studied with an enantiomeric excess up to 50% (Galman et al., 2012).

The complications of the stereoselectivity in carbon-carbon bond elongation have drawn a huge attention from a number of researches to use enzymes in asymmetric synthesis due to their naturally high stereoselectivity. In addition to transketolase and other ThDP dependent enzymes (Breuer and Hauer, 2003; Resch et al., 2011; Sprenger and Pohl, 1999), several enzymes including aldolases (Enders et al., 2005; Fessner, 1998; Machajewski and Wong, 2000; Samland and Sprenger, 2006), hydroxynitrile lyase (Breuer and Hauer, 2003) have gained interests for similar purposes. In the carbohydrate synthesis, transketolases can provide some advantages over some aldolases and these are non-requirement of dihydroxyacetone phosphate, simple purification process and monitoring due to the fact that a non-phosphorylated compound is produced (Kobori et al., 1992; Toone and Whitesides, 1991). In addition, since transketolase only accepts hydroxylated substrates with 2*R* configuration, racemic hydroxylated aldehydes can be used as a starting material (Kobori et al., 1992).

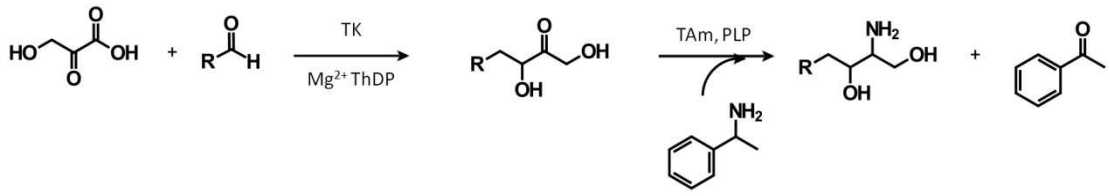
1.1.4.3.2. Substrate spectrum of TK and its potential

The product of transketolase contains a 1,3-dihydroxy ketone motif, a functional group that can be found in carbohydrates(Kobori et al., 1992), and corticosteriods (Hailes et al., 2009; Hailes et al., 2010). The substrate spectrum studies of transketolases from different species revealed transketolases can accept not only natural substrates but also non-phosphorylated sugars (Demuynck et al., 1991; Kobori et al., 1992), non-hydroxylated aldehydes, cyclic aldehydes (Hobbs et al., 1993; Morris et al., 1996), andheteroaromatics (Humphrey et al., 1995). In addition to aldehyde substrates, transketolase can also catalyse the transfer of the ketol group to nitroso aromatics (Corbett and Corbett, 1986). Mutagenesis studies of transketolase from *E. coli* have further expanded the synthetic potential of transketolase. Initially, *E. coli* transketolase libraries were found to have improved activity towards propionaldehyde (Hibbert et al., 2008) with enhanced and reverse stereoselectivity (Smith et al., 2008). Several mutants also were identified to accept cyclic and longer chain non-hydroxylated aldehydes (Cázares et al., 2010) heteroaromatics and benzaldehyde derivatives (Galman et al., 2010) with high enatiomeric excess in all cases. So far, transketolase has been used in mostly multi-step synthesis of several compounds including furaneol which is a caramel-like flavour (Hecquet et al., 1996; Hecquet et al., 1994), glycosidase inhibitors; 1,4-Dideoxy-1,4-imino-D-arabinitol (Ziegler et al., 1988) and *N*-hydroxypyrrolidine (Humphrey et al., 2000), (+)-*exo*-brevicomin (Myles et al., 1991). For single step synthesis, transketolase tends to be used for simple carbohydrate

synthesis such as ketoase sugars (Bolte et al., 1987) and deoxysugars (Hecquet et al., 1994). In addition, the production of ketose sugar can start from natural compound by deamination of serine to produce hydroxypyruvate for transketolase (Demuynck et al., 1990; Hecquet et al., 1996).

1.1.4.3.3. Transketolase and transaminase: amino alcohols

The dihydroxy ketone motif can be used as a building block for the synthesis of aminodiols which is a functionality found in several pharmaceutical compounds. The synthesis route of aminodiols could serve as a platform for the synthesis of sphingolipids (Hailes et al., 2010; Hailes et al., 2010), transition-state analogues of certain sugars which are used as glycosidase inhibitors (Takayama et al., 1997), and antibiotics chloramphenicol and derivatives (Nagabhushan et al., 2000). Although the chemosynthetic routes for the production of the aminoalcohols are available, they tend to involve the use of toxic compounds and multiple steps are required (Smith et al., 2010). The emerging of a new pathway by coupling transketolase and transaminase has shown a potential in the synthesis of optically pure aminodiols with higher atom efficient but fewer steps and complications (Hailes et al., 2010).



Scheme 1.3 The novel transketolase-transamiase pathway for the production of aminodiol via dihydroxyketone intermediate.

The novel transketolase-transaminase pathway was first used for the synthesis of 2-Amino 1,3,4 butanetriol (ABT) which can be used as a building block for Nelfinavir (Ingram et al., 2007). The wild type *E. coli* transketolase catalyses the synthesis of erythrulose from hydroxypyruvate and glycolaldehyde. Erythrulose is then aminated by β-alanine-pyruvate aminotransferase using *S*-methylbenzylamine as an amino donor. Although one pot synthesis was successful when using whole cell biocatalyst and clarified lysate, transamination showed to be a potential obstacle due to its low activity. This made the overall reaction rather slow and low yield was observed (Ingram et al., 2007). This has driven the search for a new transaminase with a better activity. ω-transaminase from *Chromobacterium violaceum* DSM30191 (CV2025) was found to accept a wide range of

substrates including aliphatic aldehydes, cyclic aldehydes, aromatic aldehydes, keto acids, and erythrulose (Kaulmann et al., 2007). In addition, a few engineered *E. coli* transketolase variants have been shown to have higher activity towards non-hydroxylated aldehyde (Hibbert et al., 2008), cyclic aldehydes (Cázares et al., 2010), and aromatic aldehydes (Galman et al., 2010). This expands the range of the products that could be synthesised by the two enzymes. Propionaldehyde, together with the best transketolase mutant D469T, was further coupled with CV2025 transaminase to explore this feasibility and to model the production process of (2*S*,3*S*)-2-aminopentane-1,3-diol (APD) (Smith et al., 2010). All the information was then integrated to design the processes and tools required for the synthesis of other2-amino-1,3-diols (Hailes et al., 2010). With the aid of process design, the coupling of transketolase wild type and mutant with CV2025 transaminase was shown to give higher yields and optical purity for both ABT and APD (Rios-Solis et al., 2011).

1.2. Research aims

The overall aim of this research is to construct transketolase mutants with higher activity towards aromatic aldehydes and integrate these variants with transaminase to produce novel chiral amino alcohols. This research consists of four major sequential steps as designated into four chapters. Each is briefly described below.

In the first chapter, two sets of libraries were constructed and the target residues were chosen from two different rationales. Their stabilities and activities towards glycolaldehyde and propionaldehyde were then compared. Eight variants including single, double, and a triple mutants from the first library were constructed by randomly combining three existing single-mutant transketolase variants that were reported to have higher activity towards glycolaldehyde than wild type. In the second library, the double mutants that were constructed and the two target residues were co-evolved residues. The information from this study will aid the design of the future library in order to compromise both activity and stability.

The second aim is to identify the factors governing the bioconversion of the aromatic aldehydes by transketolase and to find which mutants have the highest activity towards some aromatic aldehydes. Although a few transketolase mutants were identified to accept benzaldehyde and derivatives, very slow conversion rates were observed (Galman et al., 2010). Acid aromatic aldehydes were chosen to study if enzyme-substrate affinity could be the key factor. Site directed mutagenesis was employed to identify the key

mutation residues that allow transketolase to accept aromatic aldehyde which can be used as a pseudo wild type for the next library design. This strategy was also used to study how these residues control the bioconversion of aromatic aldehydes.

The third aim is to create a small transketolase library by introducing site saturation mutagenesis to certain residues within the active site. The knowledge from the combination of the single mutants in the first part will be used in designing the stable library while the literature review on the functions of certain residues would justify which residues should be targets or remained untouched. The improved mutants and the substrates of interests will be part of the novel transketolase-transaminase pathway.

Lastly, transketolase and transaminase will be coupled *in vitro* to produce a novel aromatic amino alocohol, a new derivative of chloramphenicol building block. The factors and bottlenecks associated with the process will be identified for the process improvement and strategy design.

2. Chapter 2: Materials and methods

2.1. Chemicals

Unless otherwise stated, all the chemicals were purchased from Sigma-Aldrich, Poole, UK. Lithium hydroxypyruvate was prepared from bromopyruvic acid and lithium hydroxide according to the previous method and stored at -20 °C (Morris et al., 1996).The purity of each batch was tested by HPLC according to the protocol below. Propionaldehyde was purchased from AcroSeal (Acros Organics, Fisher Scientific, UK). Standard dihydroxy ketones were synthesised by biomimetric reaction (Smith et al., 2006) except for 3-(1,3-dihydroxy-2-oxopropyl)benzoic acid and 4-(1,3-dihydroxy-2-oxopropyl)benzoic acid which were synthesised by transketolase mutants and characterised according to the publication (Payongsri et al., 2012). The standard dihydroxy ketones were kept at 4 °C.

All the buffers, liquid phase of the HPLC, and liquid media were prepared using RO water which was purified by Elix Water Purification system (Millipore, UK). This produces water with greater than 5 MΩ.cm resistivity at 25 °C, less than 30 ppb of total organic carbon (TOC), and less than 1cfu/ml of bacteria.

The water used in PCR reaction including dissolving primers was ultrapure water which was purified by Milli-Q® Advantage A10® system (Millipore, UK).

Unless otherwise stated, the pH of all the solutions wasadjusted by the addition of sodium hydroxide solution or hydrochloric acid.

2.2. Media and buffer reagent preparation

2.2.1. Ampicillin (Amp)

Ampicillin stock solution was prepared by dissolving ampicillin powder in RO water to the final concentration of 50 mg/ml. This was then filtered through sterile 0.2 µm filter.The ampicillin stock can be aliquoted into 1.5-ml sterile eppendorf tubes and stored at -20 °C. The working concentration of ampicillin in the culture media was 150 µg/ml which was achieved by adding 30 µl of this stock solution per 10 ml of the media.

2.2.2. Lysogenic broth (LB)

Lysogenic broth, known asLuria-Bertani broth, was prepared by dissolving 5 g yeast extract, 10 g tryptone, and 10 g sodium chloride in RO water with a final volume of 1 l. The pH of the mixture was adjusted to 7.0. The liquid was then autoclaved for sterilisation.

2.2.3. LB agar with 150 µg/ml ampicillin (LB-Amp)

The composition and the preparation of LB agar was the same as LB broth but select agar was also added to the final concentration of 15 g/l. After autoclave, the liquid was left to cool down to approximately 50 °C before adding ampicillin. 90 µl of this ampicillin stock solution was added to 20 ml of LB agar which was then poured into a petri dish and allowed to set. The LB-Amp plates can be stored at 4 °C for a month.

2.2.4. Tris buffer

50 mM Tris buffer was prepared by dissolving 7.88 g of Tris-HCl in 1 l ofRO water in a borosilicate bottle. The pH was then adjusted to 7.0 by the addition of 46 % (W/V) sodium hydroxide. To prevent microbial contamination and precipitation, Tris buffer was only taken out from the bottle by directly pouring into a smaller container which was then pipetted to the desired volume.

2.3. Transketolase production

All transketolase mutants with 6 his-tag at the N-terminuswere expressed under the control of its own nativepromoter in the *tktA* gene which was cloned into pQR791 plasmid(Martinez-Torres et al., 2007) which is a modification of pQR711 plasmid (French and Ward, 1995). The amino acid sequence of wild type TK is illustrated below. The positions of all amino acids mentioned throughout this thesis excluded the 6 histidine residues (underlined) used for affinity purification. The plasmid map is shown in Figure 2.1.

```
  1 MHHHHHHSSR KELANAIRAL SMDAVQKAKS GHPGAPMGMA DIAEVLWRDF LKHNPQNPSW
 61 ADRDRFVLSN GHGSMLIYSL LHLTGYDLPM EELKNFRQLH SKTPGHPEVG YTAGVETTTG
121 PLGQGIANAV GMAIAEKTLA AQFNRPGHDI VDHYTYAFMG DGCMMEGISH EVCSLAGTLK
181 LGKLIAFYDD NGISIDGHVE GWFTDDTAMR FEAYGWHVIR DIDGHDAASI KRAVEEARAV
241 TDKPSLLMCK TIIGFGSPNK AGTHDSHGAP LGDAEIALTR EQLGWKYAPF EIPSEIYAQW
301 DAKEAGQAKE SAWNEKFAAY AKAYPQEAAE FTRRMKGEMP SDFDAKAKEF IAKLQANPAK
361 IASRKASQNA IEAFGPLLPE FLGGSADLAP SNLTLWSGSK AINEDAAGNY IHYGVREFGM
421 TAIANGISLH GGFLPYTSTF LMFVEYARNA VRMAALMKQR QVMVYTHDSI GLGEDGPTHQ
481 PVEQVASLRV TPNMSTWRPC DQVESAVAWK YGVERQDGPT ALILSRQNLA QQERTEEQLA
541 NIARGGYVLK DCAGQPELIF IATGSEVELA VAAYEKLTAE GVKARVVSMS STDAFDKQDA
601 AYRESVLPKA VTARVAVEAG IADYWYKYVG LNGAIVGMTT FGESAPAELL FEEFGFTVDN
661 VVAKAKELL
```

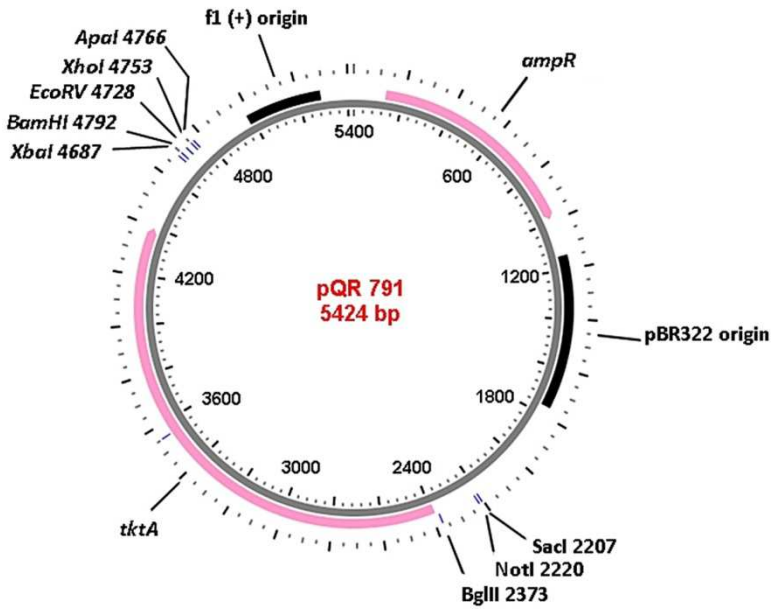


Figure 2.1 The map of pQR791 plasmid.

In the chapter 3, the *E. coli* strain used is BL21Gold(DE3) while the rest of the study, the host cell is XL10 gold (Stratagene, La Jolla, California, USA).

The glycerol stock of each mutant was streaked on an LB-Amp plate and incubated at 37 °C for 18 hours. A single colony was picked by a sterile tip and inoculated into 10 mlLB media with 150 µg/ml ampicillin in a falcon tube. The culture was then shaken at 250 rpm, 37 °C for 16-18 hours before harvested or transferred to a larger volume of fermentation. For a larger scale, 10% inoculums volume was used and the fermentation was performed at the same conditions for 6 hours before harvesting.

To harvest, the culture was aliquoted into 50-ml falcon tubes and centrifuged at 400 rpm for 10 minutes. The supernatant was discarded and the cell pellet was either used fresh or store at -20 °C.

2.4. Preparation of clarified lysate

The frozen cell paste was thawed on ice and resuspended in 2-5 ml 50 mM Tris buffer, pH 7.0. The volume of Tris buffer added depends on the culture volume and the required transketolase concentration for further use. This was then transferred to 7-ml bijou tube and lysed by sonication (MSE Soniprep 150 probe, Sanyo) on ice with 10 s on, 15 s off for 10 cycles. The mixture was then centrifuged at 17,700 g for 10 min at 4 °C. The supernatant containing transketolase was used fresh or stored at -80 °C in 150 µl aliquots for up to 1 month.

2.5. Protein concentration quantification

2.5.1. Bradford assay

The total protein concentration was measured by Bradford assay using bovine serum albumin (BSA) as a standard protein. 20 – 70 µlof 0.6 mg/ml BSA stock solution was added into the 1 ml cuvettes. 1 ml of Bradford reagent was then added to each cuvette at 15 s interval. The mixture was incubated for 5 minutes before measuring the absorbance at 595 nm. The final BSA concentration in the cuvette was then plotted against the OD595 to make a standard graph. Only the linear data points were taken. To measure the total protein concentration in the clarified lysate, 1 ml of Bradford reagent was added into a 1 ml cuvette containing 10 µl of the clarified lysate. After 5 minutes of incubation, the OD595 was read and the final concentration of the total protein in the cuvette was calculated. The actual protein concentration before the addition of Bradford reagent was then re-calculated back through the dilution factor. If the OD595 of the clarified lysate was beyond standard graph, it was diluted to fit within the range of the standard graph. All quantifications were done in duplicated.

Example

The final concentration of BSA in the cuvette when adding 1 ml of Bradford assay to 20 µl of 0.6 mg/ml BSA is

= 0.6 mg/ml * 0.020 ml / (1ml+ 0.020ml)

= 0.0118 mg/ml

2.5.2. SDS-PAGE and densitometry

After measuring the total protein concentration within the clarified lysate, the lysate was mixed with 2xLaemmli Sample Buffer (Bio-Rad Laboratories, UK) to reach the final protein concentration of 1.0 – 1.5 µg/µl. When the total protein concentration is very high, the lysate may be diluted with 50 mM Tris buffer prior mixed with 2xLaemmli Sample Buffer. The mixture was then heated at 95 °C for 10 minutes to fully denature the protein. To minimise the experimental error due to evaporation, all samples were prepared in 500 µl tube with at least 50 µl of sample. After heating, the sample was briefly centrifuged, gently mixed and briefly centrifuged again to get rid of any bubbles. 10 – 15 µl of the denatured sample was loaded into each well so that the final amount of protein per well is between 10 – 15 µg. Denatured BSA was also loaded into the same gel at the final quantity

of 1.2 – 3.6 µg per gel to make a standard correlation between the intensity of the band and the amount of protein within the band as showed in Figure 2.2. 5 µl of prestained protein ladder PageRuler™ (Fermentas, UK) was loaded into one of the well to confirm the size of transketolase if necessary.

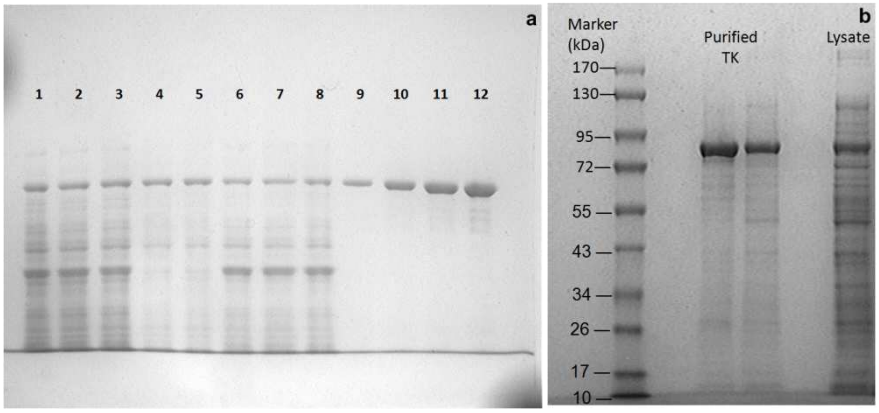


Figure 2.2 An example of SDS-PAGE gel with standard TK gradient.

a) Lanes 1-3 are whole cell sample. Lanes 3-4 are clarified lysate. Lanes 6-8 are resuspended pellets. Lanes 9 -12 are TK gradient. b) The SDS-PAGE of His-tag purified TK and clarified lysate. The molecular weights of the markers are shown accordingly.

The migration distance of each marker protein in Figure 2.2 b was plotted against the log(molecular weight kDa), a linear best fit line was obtained. The migration distances were measured when the image was 16.58 cm long and 13.5 cm wide. The top most part of the image was assigned to be the starting point of protein migration.

Table 2.1 The migration distances of the protein markers, their molecular weights, and their log(molecular weight).

MW(kDa)	Migration distance (cm)	Log (MW)
17	15.68	1.230
26	13.89	1.415
34	12.41	1.531
43	10.51	1.633
55	8.93	1.740
72	7.14	1.857
95	5.81	1.978
130	4.16	2.114
170	3.17	2.230
TK	7.54	

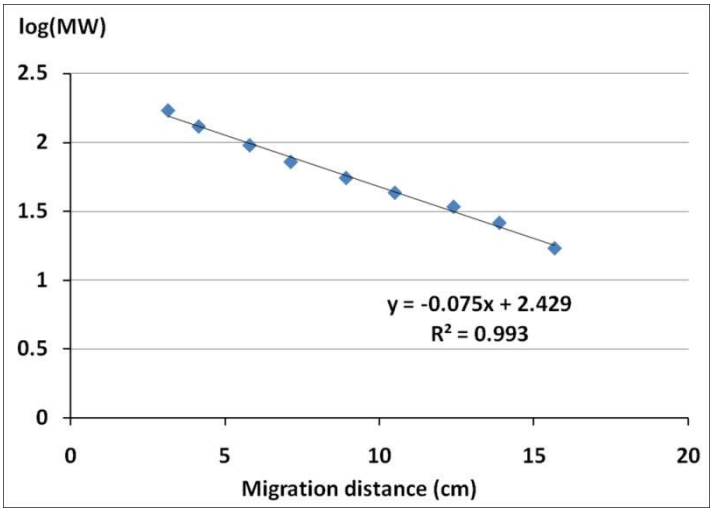


Figure 2.3 The plot between the migration distances of the protein markers and their log(MW).

The migration of transketolase in the same gel was found to be 7.54 cm. Using the linear best fit from the plot in Figure 2.3, the log(MW) of transketolase would be 1.86. Therefore, the molecular weight of transketolase determined from the SDS-PAGE gel would be 73 kDa. The molecular weight determined from the sequence of transketolase wild type in section 2.3 is 74.8 kDa.

12-well, 10 cm x 10 cm, 10% precast SDS-PAGE gels and 20x running buffer were purchased from Expedeon (Harston, UK).SDS-PAGE was performed at a constant voltage of 125 V for 100 minutes. In order to stain the gel, the gel was heated in Coomassie Brilliant Blue until boiling and left to stand for 2 minutes. The gel was then rinsed with RO water and boiled for 10 minutes in 500-ml fresh RO water to destain. Excess dye was further removed by destaining in 50% (V/V) methanol with 10% (V/V) acetic acid for a few hours. The gel was then visualised under white light and taken photo by GelDoC-IT® (UVP, UK). The intensity of each band was measured using the protocol available from Imagesoft software, LabWorks™, version 45.00.0 (UK).

In order to validate the use of BSA as a standard protein in the densitometry, the correlations between the amount of protein and the intensity of BSA and purified transketolase were compared. The concentration of his-tag purified transketolase has been quantified by measuring the absorbance at 280 nm with an extinction coefficient of 93,905 $M^{-1}cm^{-1}$ and molecular weight of 72260.82 g/mol(Aucamp et al., 2008). In Figure 2.4Figure 2.4 below, both proteins show similar correlation with less than 5% difference.

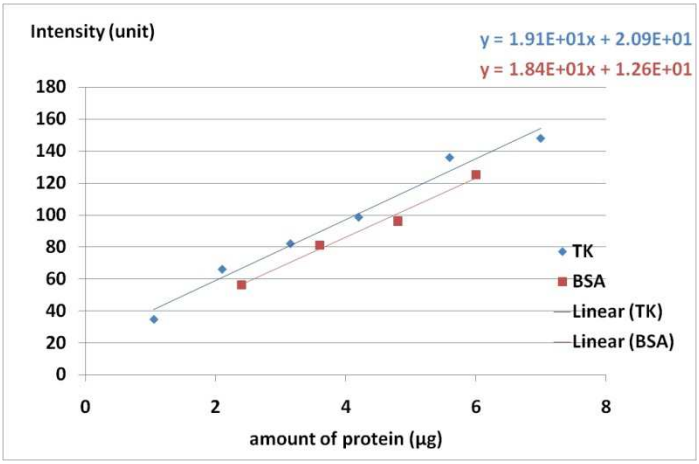


Figure 2.4 The relationship between the intensity of the band and the amount of ◆ purified TK and ■ BSA

Transketolase concentration was calculated from the amount of transketolase in the band divided by the actual volume of the lysate loaded into the well.

2.6. Enzymatic reaction

2.6.1. Cofactor solution preparation

Both ThDP and Mg^{2+} were prepared together as 10x stock solution. A solution of 24 mM ThDP, 90 mM $MgCl_2$ was prepared in 50 mM Tris buffer and the pH was adjusted to 7.0. This was kept as 150 µl aliquots in -20 °C.

2.6.2. Substrate preparations

For glycolaldehyde, 300 mM of glycolaldehyde and 300 mM HPA were separately prepared in 50 mM Tris buffer and adjusted the pH to 7.0. The two substrates were then mixed with equal volume to give 150 mM substrate stock solution which is 3x the concentration of the reaction condition. Both substrates can be stored at -20 °C in 180 µl aliquot. For other aldehyde substrates, only freshly prepared substrate solutions were used due to the risk of oxidation. 25 ml glass vials were used for the preparation of all other aldehyde substrates.

For propionaldehyde, 87 mg of propionaldehyde was added to 5 ml of 50 mM Tris buffer with 80 µl of 3 M sodium hydroxide. This adjusts the pH of the aldehyde solution to 7.0 without the need to open the container. 800 µl of 300 mM propionaldehyde was then mixed with 800 µl of 300-mM HPA in a HPCL vial which was quickly closed by a snap cap to prepare 3x substrate solution.

For 3-formylbenzoic acid and 4-formylbenzoic acid, their solubilities are lower than glycolaldehyde and propionaldehyde, so they were prepared at 150 mM and adjusted the pH to 7.0. This was then used to dissolve HPA to get the final concentration of 150 mM each.

2.6.3. Activity assay

All reactions were run in triplicate at the final volume of 300 µl at 22 °C. Prior starting the reaction, 30 µl of clarified lysated was incubated with 30 µl of 10X cofactor and 140 µl of 50 mM Tris buffer, pH 7.0 which makes the total volume of 200 µl. After 20 minutes of incubation, 100 µl of 3X substrate stock solution was added to the enzyme-cofactor solution to start the reaction at 15 s interval. 20 µl of reaction sample was taken at certain time points and quenched in 180 µl 0.1% TFA for glycolaldehyde and propionaldehyde. For aromatic substrate, 20 µl samples were quenched in 380 µl 0.1% TFA. All samples were quenched in 500-µl reaction tubes which were centrifuged at 13,000 rpm for 3 minutes to remove any precipitate. The supernatants were analysed by HPLC to quantify the substrate and product concentration. The reaction of glycolaldehyde was performed in polystyrene 96-well plate which was covered by its lid. For other aldehydes, the reactions were performed in 1.5-ml HPLC vials which were closed by snap caps.

2.7. HPLC system

Two separate HPLC systems were employed in this study for different types of compound separation. Both HPLCs were controlled by Chromeleon software (Dionex, UK). The chromatograms and peak analysis were also processed using the same software.

2.7.1. HPLC system for aliphatic compounds

At least 170 µl of the supernatants was loaded onto a polystyrene 96 well plate and covered with aluminium foil to prevent evaporation. The sample was taken by FAMOS autosampler (Dionex, UK) which was then injected into Aminex HPx-87H, 300x7.8mm column (Bio-Rad, UK) which was maintained at 60 °C. Micro-Guard Cation H Refill Cartridges, 30 x 4.6 mm (Bio-Rads, UK) was also used to prevent column contamination. The separation was performed by isocratic flow of 0.1% TFA in water at the flow rate of 0.6 ml/min which was maintained by Ultimate 3000 isocratic analytical pump (Dionex, UK). The UV absorption at 210 nm was monitored at the data collection rate of 5Hz by AD20 absorbance detector (Dionex, UK) for the detection of HPA, erythrulose, 1,3 dihydroxy

pentan-2one which have the retention time of 8.4, 11.5, and 15.2 minutes, respectively. For glycolaldehyde reaction, the HPLC assay time was 15 minutes whereas the propionaldehyde is 20 minutes. The 0.1 % TFA mobile phase was degassed by sparging with helium for 10 minutes before use.

2.7.2. HPLC system for aromatic compounds

360 µl of the supernatant was transferred into 1.5-ml HPLC vials and closed with crimp caps. The sample was taken by Ultimate 3000 autosampler (Dionex, UK) and injected into ACE5 C18 reverse phase column (150 × 4.6 mm) which was maintained at 30 °C. The mobile phase was pumped in at the flow rate of 1.0 ml/min starting with 85% of 0.1% TFA and 15% of 100% acetonitrile (CHROMASOLV® HPLC gradient grade) which was linearly increased to 72% in 9 minutes. This followed by 2-minute equilibration period by pumping 85% of 0.1% TFA and 15% acetonitrile. UV detection at 210, 250, and 275 nm were simultaneously used throughout the process.

For 3-hydroxybenzaldehyde, the same column was used but the mobile phases were changed to 0.2 M acetic acid and 80% (V/V) methanol (in water). The mobile phase starts with constant flow of 90% 0.2 M acetic acid and 10% of 80% methanol for 5 minutes. This is followed by a linear increase of 80% methanol from 10% to 60% in the 14^{th} minute which was maintained for another 6 minutes. Then the column was equilibrated for 3 minutes by 90% 0.2M acetic acid and 10% 80% (V/V) methanol.

2.8. Standard product and substrate graphs

The commercially available standard substrates and products were purchased from SigmaAldrich. The racemic aromatic dihydroxyketones were synthesised by biomimetic reaction (Smith et al., 2006) which have been purified and characterised by NMR (David Steadman, Chemistry Department, UCL). All reagents were dissolved in 0.1% TFA.

For HPA and erythrulose which were analysed by the first HPLC system, their final concentrations in TFA were between 0.05 – 5.0 mM. This range represents 0.5 – 50 mM concentration before the reaction samples were diluted 10 times upon quenching in 0.1% TFA. For aromatic aldehydes, the final concentrations of the aldehydes were between 0.05 – 2.5 mM which represents 1 – 50 mM of aldehydes before 20 times dilution. At least five concentration points were prepared and the R^2 value of the linear standard graph must be greater than 0.96.

2.9. Initial rate and specific activity calculation

The graph of product versus time was plotted in Microsoft Excel 2007. If the data are linear, linear best-fit line was used to find the initial rate.

$$\text{Specific activity } (\mu\text{mol mg}^{-1}\text{ min}^{-1}) = \frac{\text{initial rate } (\frac{\text{mM}}{\text{min}})}{\text{enzyme concentration } (\frac{\text{mg}}{\text{ml}})}$$

If the data are not in the linear range, the whole data set was analysed by SigmaPlot 12. The exponential rise to maximum; $y = a\left(1-e^{-b\,x}\right)$ was fitted into the data and SigmaPlot then predicted the equation that fits the data. Y is product concentration and x is time. SigmaPlot also returns the predicted values of the product concentration at time zero until the last data point. The initial values of product concentration and time were then used to calculate the initial rate.

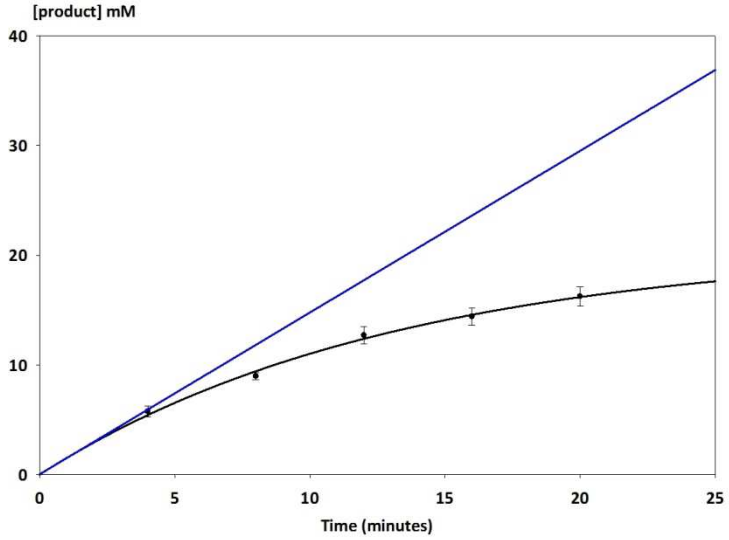


Figure 2.5 The calculation of initial rate using SigmaPlot.

The raw data are shown as ● and the calculated exponential rise to the maximum fitted the data is showed in black line. The linear range at the beginning of the reaction time was calculated which was then used to calculate the linear best fit as showed in blue line.

2.10. Synergy calculation

The synergetic effect upon multiple mutations was calculated based on the changes of free energy calculation relative to wild type or pseudo wild type (Mildvan et al., 1992; Schreiber and Fersht, 1995; Wells, 1990).

$$\Delta\Delta G_{(x,y)} = \Delta\Delta G_{(x)} + \Delta\Delta G_{(y)} + \Delta G_{(I)}$$

Or $\Delta G_{(I)} = \Delta\Delta G_{(x,y)} - (\Delta\Delta G_{(x)} + \Delta\Delta G_{(y)})$

$\Delta\Delta G_{(x)} = RT \ln (A_x/A_{WT})$; the change in the free energy from wild type due to the mutation at site x

$\Delta\Delta G_{(y)} = RT \ln (A_y/A_{WT})$; the change in the free energy from wild type due to the mutation at site y

$\Delta\Delta G_{(x,y)} = RT \ln (A_{x,y}/A_{WT})$; the change in the free energy from wild type due to the double mutation at site x and y

$\Delta G_{(I)}$ is the coupling energy

This could be simplified as $\Delta G_{(I)} = RT\{ \ln (A_{x,y}/A_{WT}) - \ln (A_x/A_{WT}) - \ln(A_y/A_{WT})\}$

$$\Delta G_{(I)} = RT \ln \{(A_{x,y}\, A_{WT})/(A_x A_y)\}$$

A is the kinetic parameter such as specific activity, K_M, and k_{cat}.

R is gas constant; 8.3144621J K^{-1} mol^{-1}

T is temperature in K.

The value of $\Delta G_{(I)}$ determines the effects of the double mutation whether both sites are additive, partially additive, neutral, negative, or epistatic (antagonistic) (Mildvan et al., 1992).

If $\Delta G_{(I)} \approx 0$, both mutants have additive effect

If $\Delta G_{(I)} < -\Delta G_{(y)}$, the second mutant exerts epistatic effect on the first on

If $\Delta G_{(I)} = -\Delta G_{(y)}$, no synergy between x and y

If $-\Delta G_{(y)} < \Delta G_{(I)} < 0$, the second mutant has partially positive effect on the first one

Since all the ΔG terms have RT terms, the $(A_{x,y}\, A_{WT})/(A_x A_y)$ and A_x or A_y can be compared directly as well.

2.11. *E. coli* strain and glycerol stock preparation

The initial BL21Gold DE3 carrying wild type *tktA* gene in pQR791 plasmid was provided by Leonardo Rios. The D469T mutant, also in pQR791 plasmid but expressed in XL10 Gold strain, was provided by Phattaraporn Morris. Both were streaked on LB-Amp

plates to get single colonies which were used to inoculate in 10 ml LB broth with 150 µg/ml ampicillin. They were cultured overnight according to the above protocol. 4 ml of the cultured was mixed with 4 ml of 40 % filtered-sterile glycerol which was then stored at -80 °C in 1 ml aliquots in sterile eppendorf tubes.

2.12. Plasmid extraction and storage

A single colony was picked and cultured in 10 ml LB media with 150 µg/ml ampicillin as above protocol. 1.5 ml of the culture was centrifuged at 17,000 g for 5 minutes in 1.5 ml eppendorf tube. The supernatant was discarded and another 1.5 ml of the culture was added to the same tube which was centrifuged for another round. This cell paste from 3 ml of the culture was used for plasmid extraction by QIAprep Spin Miniprep Kit (QIAGEN, UK). The plasmid was eluted from the column by Milli-Q® water. The unused pellets and the extracted plasmid were stored in -20 °C. The integrity of the plasmid was checked by DNA gel electrophoresis.

2.13. DNA gel electrophoresis

0.7 % (W/V) agarose gel was used throughout the study. 0.21 g of agarose and 30 ml of TAE buffer were added into a 100 ml flask which was then heated in the microwave until boiling and all agarose was dissolved. 5 µl of 10,000X Sybr® safe (Invitrogen, UK) was added while the mixture was still hot. Once the gel was cooled down to approximately 50 °C, it was poured into the casting tray with the comb inserted and sealed by silicone gaskets. Once the gel was set, the gaskets and the comb were removed. The tray was transferred to the horizontal gel tank where the TAE buffer was poured into until the gel was submerged.

5 µl of DNA samples was mixed with 5 µl 6x DNA loading buffer (Merck, Nottingham, UK). 8 µl of the mixture was loaded into each well. 5 µl Hyperladder I (Bioline, UK) was loaded onto the gel to be used as DNA marker. The gel was run at a constant voltage of 70 volt for 45 minutes. The gel was visualised under the UV light by GelDoC-IT® (UVP, UK).

2.14. Transketolase library construction

2.14.1. Primer design and preparation

All the primers were designed to have two GC at the 3’ end with the length between 28-33 bases. The base pairs to be replaced are at the middle of the primer. The codon with high usage and least substitution was selected for a single amino acid replacement. For random mutation, NHS was used. All the primers are listed in Table 2.2 below.

Table 2.2 The sequences of all the primers used.

H26Y Fwd	CAGAAAGCCAAATCCGGT**T**AC**C**CGGGTGCCCCTATGG
H26Y Rev	CCATAGGGGCACCCGGGTAACCGGATTTGGCTTTCTG
H461S Fwd	GGTGATGGTTTACACC**AG**C**C**GACTCCATCGGTCTGG
H461S Rev	CCAGACCGATGGAGTCGCTGGTGTAAACCATCACC
D469T Fwd	TCGGTCTGGGCGAA**AC**C**C**GGGCCGACTCACCAG
D469T Rev	CTGGTGAGTCGGCCCGGTTTCGCCCAGACCGA
D469Y Fwd	TCGGTCTGGGCGAA**T**AC**C**GGGCCGACTCACCAG
D469Y Rev	CTGGTGAGTCGGCCCGTATTCGCCCAGACCGA
R520V Fwd	CACTGATCCTCTCC**GT**T**C**AGAACCTGGCGCAG
R520V Rev	CTGCGCCAGGTTCTGAACGGAGAGGATCAGTG
A29E Fwd	GGTCACCCGGGT**G**A**A**CCTATGGGTATGGC
A29E Rev	GCCATACCCATAGGTTCACCCGGGTGACC
F434A Fwd	CGTACACCTCCACC**GC**C**C**TGATGTTCGTGG
F434A Rev	CCACGAACATCAGGGCGGTGGAGGTGTACG
R520Q Fwd	GATCCTCTCC**CAG**CAGAACCTGGCGCAG
R520Q Rev	CTGCGCCAGGTTCTGCTGGGAGAGGATC
R358E/D Fwd	GAAAATCGCCAGC**GAN**AAAGCGTCTCAGAATG
R358E/D Rev	CATTCTGAGACGCTTTNTCGCTGGCGATTTTC
R358X Fwd	GAAAATCGCCAGC**NHN**AAAGCGTCTCAGAATG
R358X Rev	CATTCTGAGACGCTTTNDNGCTGGCGATTTTC
S385X Fwd	GCTGACCTGGCGCCG**NHS**AACCTGACCCTGTGG
S385X Rev	CCACAGGGTCAGGTT**SDN**CGGCGCCAGGTCAGC
D469E Fwd	TCGGTCTGGGCGAA**GA**A**A**GGGCCGACTCACCAG
D469E Rev	CTGGTGAGTCGGCCCTTCTTCGCCCAGACCGA

Fwd and Rev refer to forward and reverse primers respectively. The target codons are underlined and the nucleotides that have been changed are in bold letters.

All primers were ordered from eurofins mwg/operon (UK) as lyophilised form, salt free grade. The primers were briefly centrifuged before resuspended in Milli-Q® water to the final concentration of 100 µM. This was then diluted to 5 µM at the final volume of 200 µl to be used as a working stock solution. All the primer solutions were stored at -20 °C. All primers were prepared in sterile eppendorf tubes.

2.14.2. Polymerase Chain Reaction Conditions

All the mutations were introduced to the DNA template by site directed mutagenesis kit (Stratagene, Cheshire, UK). A 50-µl PCR reaction consisted of 2 µl of forward primer, 2 µl of reverse primer, 5 µl of 10X reaction buffer, 3 µl Quick solution, 1 µl dNTP mix, 1 µl DNA template, and 36 µl of Milli-Q® water. After a brief mixing, 1 µl of DNA polymerase was added. The PCR tube was transferred to the thermal cycler(Techgene, version 13, block 20 x 0.5 ml). The details of the PCR cycle are illustrated in Table 2.3.

Table 2.3 The conditions of the PCR reaction.

Segment	number of cycle	Step	Temperature (°C)	Time
1	1	Initial denaturation	95	30 s
2	25	Denaturation	95	30 s
		Annealing	55	1 min
		Elongation	68	15 min
3	Hold		4	

For double mutant construction, two sets of primers were added to the reaction volume at the same time but the 1 µl of forward and reverse primer of each primer set were added in order to make the total concentration of all the forward and reverse primer constant. For the triple mutant, 0.7 µl was used.

All the PCR products were checked by DNA gel electrophoresis before any further use. After confirming that PCR was successful and gave the right product size, 1 µl of DpnI enzyme was added to 45 µl of PCR reaction mixture which was then left at 37 °C for 1 hour to remove the DNA template. After that, the sample can be stored at -20 °C or used straight away.

2.14.3. Transformation

1 µl of the DpnI treated PCR reaction was added to a sterile eppendorf tube which was kept on ice. The competent cell was taken from the -80 °C and left on ice for 5 minutes. 30 µl of the competent cell was added to the pre-chilled eppendorf containing the DNA and left for 20 minutes. The competent cells were subject to heat shock at 42 °C for 45 s and brought back to ice for 3 minutes. 1 ml of S.O.C media (Invitrogen, UK) was added to the competent cells which were then incubated at 37 °C, shaking at 250 rpm for 1 hour. 150 µl

of the mixture was spread on LB-Amp plate and the rest was spread onto another plate. These plates were pre dried in the incubator at 37 °C for 1 hour prior spreading the cells. These plates were left to dry then incubated at 37 °C for 18 hours. At least 6 colonies were picked to prepare glycerol stock and extract plasmid to confirm the mutation by DNA sequencing of which the service was provided by Wolfson Institute for Biomedical Research and the UCL Cancer Institute, UK.

2.15. Purified enzyme preparation

2.15.1. Buffers used in purification

2.15.1.1. Equilibration and wash buffers

The equilibration and wash buffers consist of 50 mM sodium phosphate monobasic (NaH2PO4), 300 mM sodium chloride, and 10 mM imidazole. The solution was prepared dissolving 6 g of NaH2PO4, 17.5 g of sodium chloride, and 0.68 g of imidazole in 1 l of RO water and the pH was adjusted to 8.0 by addition of 46 % (W/V) sodium hydroxide.

2.15.1.2. Elution buffer

The elution buffer consists of 50 mM sodium phosphate monobasic (NaH_2PO_4), 300 mM sodium chloride, and 250 mM imidazole. This was prepared by dissolving 6 g of NaH_2PO_4, 17.5 g of sodium chloride, and 17 g of imidazole in 1 l of RO and adjusting the pH to 8.0.

2.15.2. Purification procedure

Both transketolase and transaminase enzymes contain 6 histidine residues at their N-termini for affinity purification using His-select™ gel (Sigma-Aldrich, UK). *E. coli* strains producing either transketolase or transaminase were cultured according to the standard procedure at the fermentation volume of 200 ml (for TK see section 2.3 and for transaminase see section 6.2.3). Cell pasted obtained from 50 ml culture was resuspended in the appropriate buffer prior the preparation of clarified lysate. For the cell paste for transketolase, 5 ml of 50 mM Tris buffer, pH 7.0 was used while the cell paste for transaminase was resuspended in 5 ml of 50 mM Tris buffer, pH 7.5 with 0.4 mM PLP. Lysate was used fresh and must be kept on ice at all time.

The nickel beads were stored in 30% ethanol solution and must be resuspended in its own container prior to transferring 5 ml of these fully mixed beads to a 10-ml centrifugal column (Thermo Scientific Pierce, UK) which is fitted with a falcon tube to collect the flow-through. The column was spun at 1000 rpm for 5 minutes to remove ethanol which was collected in the falcon tube. 10 ml of equilibration buffer was added to the centrifugal column and the column was inverted a few times to mix the beads. This was then centrifuged at 1000 rpm for 5 minutes and the flow-through buffer was discarded. After this step was repeated three times, 5-ml clarified lysate was then transferred to the beads. The column was closed and mixed by attaching the column to a Rotospin Test Tube Rotator and spun at 40 rpm for 30 minutes. This binding procedure was performed in cold cabinet to maintain the temperature at 4 °C. The column was spun at 1000 rpm for 5 minutes and the flow through was discarded. 10 ml of wash buffer was added to the beads and the column was mixed by Test Tube Rotator as above for 5 minutes. The column was then spun at 1000 rpm for 5 minutes and the flow-through was discarded. This was repeated twice more times to remove nonspecific binding. 5 ml of elution buffer was then added to the beads and mixed by Test Tube Rotator for 15 minutes. The column was then spun at 1000 rpm for 5 minutes. The flow-through containing purified enzyme was collected for dialysis. The beads were resuspended in 10 ml of elution buffer for two more times but the flow through was discarded. The beads were washed with RO water and stored in 30% ethanol at 4 °C. The used beads were only used with the same enzyme variant.

The collected flow-through containing purified enzyme was transferred to dialysis tube with a molecular weight cut-off of 10 kDa (Perbio Science UK Ltd., Northumberland, UK). The tube was sealed and dialysed in 4 l of 50 mM Tris bufferat 4°C for 18 hours. For transketolase, the pH of the buffer was 7.0 while the pH for transaminase was 7.5 and the buffer also contains 0.4 mM PLP. The buffer was gently mixed by a magnetic stirrer throughout the dialysis. After 18 hours, the sample in the dialysis tube was transferred to a clean tube and stored at 4°C to be used within 7 days. The purity was assessed by SDS-PAGE gel as shown in Figure 2.2b.

3. Chapter 3: Recombination of existing single mutants

3.1. Introduction

Protein engineering is a powerful tool for improving properties of an enzyme. In order to expand an enzyme substrate spectrum using a more efficient small library size, mutations could be focussed around the active site of an enzyme (Bougioukou et al., 2009; Morley and Kazlauskas, 2005; Paramesvaran et al., 2009; Reetz et al., 2005; Toscano et al., 2007). Mutations further away from the active site could also cause conformational changes within an enzyme which result in reshaping of the active site and improving activity (Lee and Goodey, 2011; Wu et al., 2010). When considering the use of enzymes in industrial processes where organic solvents, elevated temperature, or non-physiological pH may be required, catalytic activity is not the only important parameter, as enzyme stability must also be considered.

The effects of environmental factors on the stability of the wild type *E. coli* transketolase have been investigated (Dalby et al., 2007; Jahromi et al., 2011; Martinez-Torres et al., 2007). At the same time, transketolase has been engineered to accept a wider range of substrates. In recent years, directed evolution of transketolase has led to several mutants with improved activity towards glycolaldehyde (Hibbert et al., 2007), and propionaldehyde (Hibbert et al., 2008). The D469X library has gained a particular interest due to the fact that some variants have also yielded higher and reversed stereoselectivity (Smith et al., 2008), and accepted a wider range of aldehydes including cyclic, longer chain (Cázares et al., 2010) and aromatic aldehydes (Galman et al., 2010). Besides that, the enzyme kinetic study of the D469T mutant revealed that substituting the negatively charged residue with a hydrophobic residue not only enhanced the affinity towards propionaldehyde but also significantly improved the k_{cat} (Hibbert et al., 2008).

Having showed several synthetic potential, the substrate spectrum and activity of transketolase could be further improved, possibly by introducing multiple mutations. However, the stability of the mutant has to be considered and so far, only the wild type stability has been documented. The stability of a protein can be determined from *in vitro* studies of the kinetics and the thermodynamic stability (Tokuriki and Tawfik, 2009). Since the functional protein is in a folded form, thermostability assays usually involve monitoring folding/unfolding and aggregation. Previous *in vitro* studies suggested that the more thermostable enzymes unfold and aggregate at higher temperatures (Wang et al., 2010).

Aggregation could also be observed within a cell as inclusion bodies. Previous *in vivo* studies of destabilising mutations suggested that they can influence the formation of inclusion bodies and reduce the amount of the soluble fraction within the cell (Calloni et al., 2005; Tokuriki and Tawfik, 2009; Wetzel, 1994). These stabilities, however, have been observed to trade off with the enzyme activity such that when the activity is improved, stability tends to decline (Arnold et al., 2001; Dalby, 2011; Tokuriki et al., 2008). This is because there is compensation between structural rigidity and flexibility, where a rigid structure makes an enzyme more stable, whereas flexibility is required for an enzyme to function (Arnold et al., 2001). This suggests it is best to use a more stable variant as a template for directed evolution as it can accumulate more mutations (Bloom et al., 2006; Tokuriki and Tawfik, 2009). Additional stability can be brought to the target protein through introducing consensus sequence (Lehmann and Wyss, 2001) or global suppressors (Brown et al., 2010).

Another approach to target mutations with potentially less deleterious impact is to use the evolutionary information of the proteins. Multiple sequence alignments of enzymes within the same familycan reveal that that certain residues mutate in a manner that is dependent on mutations at other residues, to maintain a structure compatible for a new function (Göbel et al., 1994; Neher, 1994). This pattern of mutation forms a network which has been expressed as statistically coupled residues (Lockless and Ranganathan, 1999) and their coupling energy (Süel et al., 2003). It was suggested that this network may be important for maintaining the stability of the folded state of a protein (Göbel et al., 1994; Socolich et al., 2005) and also the long-range residue-residue interaction (Lockless and Ranganathan, 1999).

In this study, I carried out kinetic analyses and mutagenesis work as part of a collaborative project that initially aimed to create a highly active mutant towards glycolaldehyde and propionaldehyde by recombining all the single mutants previously identified to have enhanced activities towards both substrates (Strafford et al., 2012). Preliminary work by Ed Hibbert had created several double mutants and found that all of them lost activity as well as soluble expression of the enzyme through destabilisation. I created a complementary series of rationally designed recombinant mutants and characterised them in this Chapter. Another group of double mutants were also created by John Strafford using saturation mutagenesis at two or three residues within a network of residues that were found to co-evolve in TK. The combinations between D469 and R520

were the most beneficial for stability. I also carried out repeats and additional characterisation of some of these mutants as presented here. These studies were all combined into a single paper (Strafford et al., 2012), the overall results of which will be summarised here.

3.2. Materials and method

3.2.1. Materials

The sources of all chemicals are listed in chapter 2. The bottle of propionaldehyde was closed with septum cap. The liquid was drawn out using syringe and needle. The bottle was then filled with 100% nitrogen to prevent oxidation.

3.2.2. Mutant construction

The primers used in this study were listed in Table 2.2, chapter 2 and confirmed by DNA sequencing. These residues were selected because they are either highly conserved residues in close proximity to the active centre or and medium range residues that are phylogenetic variants (Hibbert et al., 2008, 2007). All the plasmids bearing mutations were transformed into BL21Gold(DE3)because it was suggested to minimise proteolysis of the recombinant protein (Miller et al., 2007). In addition, this strain was also chosen for the expression of both transketolase and transaminase (Ingram et al., 2007; Rios-Solis et al., 2011). Since the highly active mutants would be co-expressed with transaminase for the coupling reaction, this strain was chosen.

3.2.3. Protein quantification

Total protein concentration was quantified by Bradford assay according to the protocol. The concentration of transketolase within the clarified lysate was determined from SDS-PAGE and densitometry.

3.2.4. Expression level study

Single colonies of each mutant were inoculated into 10 ml LB broth containing 150 µg/ml ampicillin. They were cultured for 18 hours at 37 °C, shaking at 250 rpm. Cells were harvested by centrifugation at 4000 rpm, 4 °C, for 10 minutes. The supernatant was discarded and the pellets were resuspened in 2 ml 50 mM Tris buffer, pH 7.0. This was split into 2 portions; 1 ml was kept as whole cell sample, another 1 ml was sonicated to analyse

the soluble and insoluble fraction. The 1 ml of the resuspended cells was transferred into 2-ml tubes and sonicated on ice for 10 cycles with 10 s on, 15 s off. They were then centrifuged at 17,000 g, 4°C, for 10 minutes. The supernatants were transferred into clean 2-ml eppendorf tubes. All the pellets were fully resuspended in 1 ml 50 mM Tris buffer, pH 7.0.

The 50 µl of whole cell, lysate, and resuspended pellet were mixed with 50 µl 2X Laemmli buffer and heat to 95 °C for 10 minutes. Up to this point, the lysate, and the pellets always had the same final volume as the whole cell sample. The total protein concentrations in the clarified lysate were determined by Bradford assay to adjust the volume of lysate to be loaded into the SDS-PAGE gel. The same volumes were applied for the whole cell and resuspended pellets. Different amount of purified transketolase were added into all the gels to make a calibration curve for each individual gel. The gels were analysed by the protocol in materials and methods. Expression level refers to the percentage concentration of transketolase in the lysate as defined below.

$$\text{Percentage concentration} = \frac{[\text{transketolase concentration in lysate}]}{[\text{total protein concentration in lysate}]} \times 100$$

$$\text{Soluble :inclusion body ratio} = \frac{[\text{transketolase in lysate}]}{[\text{transketolase in debris pellet}]}$$

3.2.5. Identification of the protein in the insoluble fraction

The insoluble fraction was prepared for SDS-PAGE analysis according to the protocol above. Protein ladder was also added to the gel. The plot of log(molecular weight) versus the migration distance of the marker was used to construct a calibration curve to identify the molecular weight of P2 protein. The band was then cut from the gel and sent for protein sequencing (Alta Bioscience, University of Birmingham, UK).

3.2.6. Activity assay

Clarified lysate was prepared according to the protocol. The 30 µl of lysate was incubated with 30 µl of 10xcofactor solution (24 mM ThDP, 90 mM $MgCl_2$) and 140 µl 50 mM Tris buffer, pH 7.0 at room temperature for 20 minutes. 3X substrate solution (150 mM HPA, 150 mM aldehyde) were prepared in 50 mM Tris buffer and adjusted pH to 7.0. The reaction was started by adding 100 µl of 3x substrate solution to the enzyme-cofactor solution. All reactions were done in triplicate. For the glycolaldehyde activity study, the

reaction was run in polystyrene, round-bottom microtitre plate and covered with a lid during the assay. The samples were taken at every 1 minute. All the propionaldehyde reactions were run in glass vials and covered with snap caps. The samples were taken at every 10 minutes for the variants containing D469T/Y for 1 hour. For other mutants, the samples were taken at every 60 minutes for 5 hours.

In order to directly compare the effect of mutation upon the substrate preference between glycolaldehyde and propionaldehyde, the same batch of clarified lysate were used to identify the activities of all the mutants towards both substrates. This eliminated any possible error from protein concentration quantification. Only freshly prepared propionaldehyde was used in all experiments to avoid substrate oxidation.

3.2.7. HPLC analysis

HPA, erythrulose, and 1,3-dihydroxypentane-2one (DHP) were quantified by HPLC using Aminex HPx-87H, 300x7.8mm column (Bio-Rad, UK) according to the protocol in chapter 2, section 2.7.1.

3.2.7.1. Standard graphs

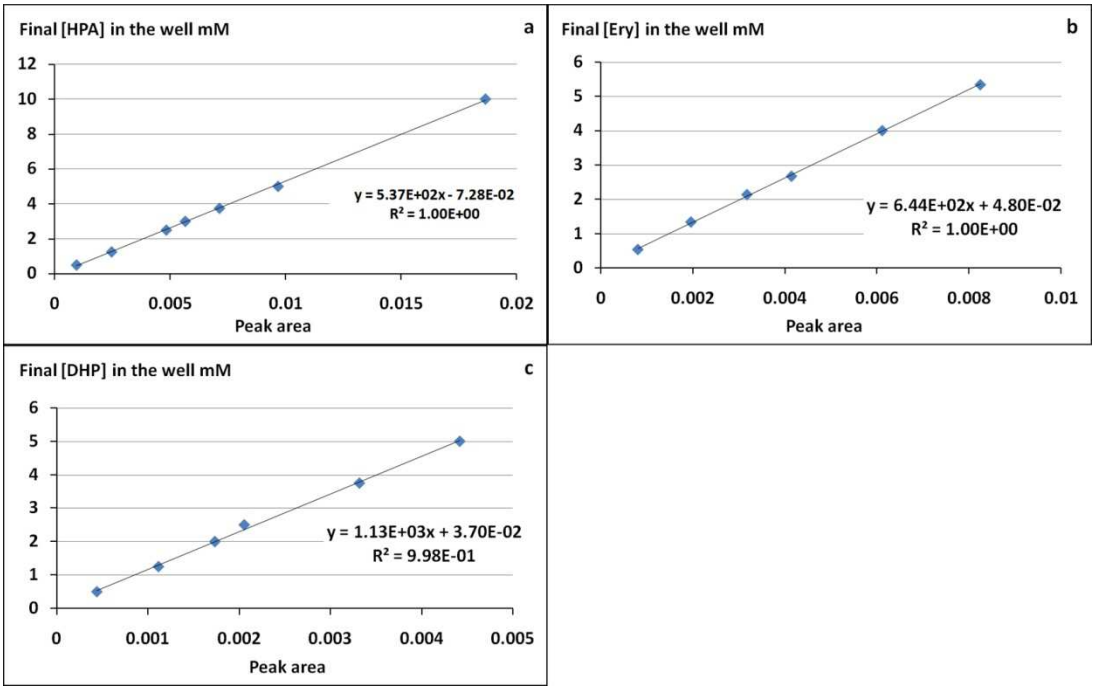


Figure 3.1 Standard graphs of a) HPA, b) erythrulose, and c) DHP.

The standard graphs were plotted based on the final concentrations of the solutions in the wells.

3.3. Results and discussion

3.3.1. The analysis of the reaction samples by HPLC

The reaction compositions can be clearly separated by the Aminex HPx-87H, 300x7.8mm column (BioRad) with the isocratic flow of 0.1% TFA as illustrated in Figure 3.2. The retention times of HPA, erythrulose, and DHP were8.4, 11.5, and 15.2 minutes, respectively. Glycolaldehyde and propionaldehyde cannot be detected by the UV absorption at 210 nm.

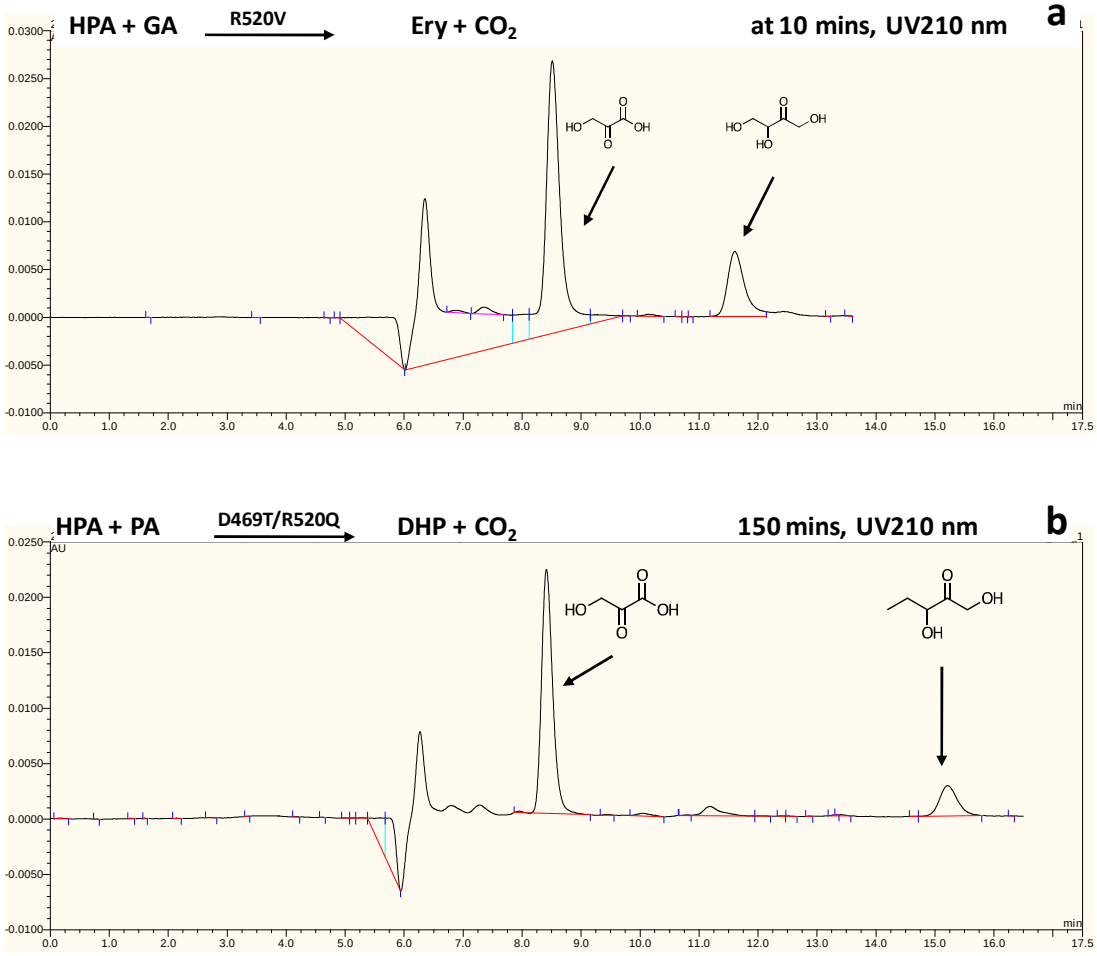


Figure 3.2 The chromatograms of reaction samples analysed by Aminex HPx-87H, 300x7.8mm column (Bio-Rad, UK) according to the protocol in Chapter 2, section 2.7.1.

The peaks are labelled with the structures of compounds corresponding to their retention times. a) the sample from HPA+GA reaction catalysed by R520V mutant. b) sample from HPA + PA reaction catalysed by D469Y.

3.3.2. The specific activities of transketolase mutants towards glycolaldehyde and propionaldehyde

The previous work illustrated that the wild type transketolase from the host strain has very low activity compared with the highly expressed wild type and this background can be neglected (Miller et al., 2007). In addition, two mutants in this study, H26Y and H26Y/D469Y, were found to be inactive. When assessing the activities of these mutants towards both glycolaldehyde and propionaldehyde, there were no background activities of the wild type enzyme from the host strain.

The mutants in this study were separated into two groups. The first group are those that have been studied with glycolaldehyde in the previous section and the mutants in the second group are those that were reported to enhance activities towards propionaldehyde (Hibbert et al., 2008) or influence the stereoselectivity of the enzyme (Smith et al., 2008). The specific activities of all the single and combined mutants towards glycolaldehyde and propionaldehyde are listed in Table 3.1. below. The activities of the wild type and mutants previously identified to have higher activities towards glycolaldehyde (Hibbert et al., 2007) and propionaldehyde (Hibbert et al., 2008) were repeated in this study and added for comparison. Both studies showed that D469T and D469Y had improved activities towards propionaldehyde although the percentage increase from wild type still differed. The double mutants containing D469T/Y also had higher activities than wild type. Since the previous report suggested that there was a strong shift in substrate preference upon the mutation at H26 and D469, the two groups will be discussed separately (Hibbert et al., 2008).

From Table 3.1, it can be seen that the specific activity of the wild type in this study was significantly higher than what was determined by Hibbert et al., (2007). Yet, the mutants identified with enhaced activity towards glycolaldehyde were found to have either similar (A29E, H461S) or lower activity (R520V) than wild type. Therefore, the experimental conditions were compared in Table 3.2 to identify the sources of deviation.

Table 3.1 The specific activities of all the mutants towards glycolaldehyde and propionaldehyde determined in this study and previously determined.

Mutant	Glycolaldehyde activity		Propionaldehyde activity		Rate GA:PA (from this study only)
	This study	Previous study [b]	This study	Previous study [c]	
	µmol mg^{-1} min^{-1}	µmol mg^{-1} min^{-1}	µmol mg^{-1} min^{-1}	µmol mg^{-1} min^{-1}	
Group I					
WT	36 (1.1)	0.65 (0.04)	0.20	0.029 (0.001)	183
A29E	38 (6.2)	1.95 (0.17)	0.16	0.10 (0.02)	232
H461S	32 (11)	3.14 (0.54)	0.14	0.040 (0.007)	223
R520V	22 (3.6)	2.3 (0.5)	0.11	0.14(0.05)	206
A29E H461S	28 (4.2)		0.2		142
A29E R520V	34 (11)		0.14		249
H461S R520V	15 (4.2)		0[a]		n.d.
A29E H461S R520V	37		0.17		224
Group II					
H26Y	0[a]		0[a]		n.d.
D469Y	n.d		2.4	0.127 (0.01)	n.d.
H26Y D469Y	0[a]		0[a]		
D469Y R520V	n.d.		1.5		
D469T			1.4	0.14 (0.001)	n.d.
D469T R520Q			1.9		
D469Y R520Q			1.2		
A29E D469Y			0[a]		
R520Q				0.69[d]	

[a] No activity was observed, [b](Hibbert et al., 2007), [c](Hibbert et al., 2008), [d](Strafford et al., 2012)

Table 3.2 The comparison of the experimental outcomes and conditions between this study and the previous study when using glycolaldehyde as a substrate.

	This study			Previously determined [b]		
Mutant	Activity μmol mg^{-1} min^{-1}	Initial rate [a] mM min^{-1}	[TK] mg/ml	Activity μmol mg^{-1} min^{-1}	Initial rate mM min^{-1}	[TK] mg/ml
WT	36 (1.1)	6.4 (0.68)	0.18	0.65 (0.04)	0.29 (0.01)	0.44
A29E	38 (6.2)	4.56 (0.18)	0.11	1.95 (0.17)	0.66 (0.01)	0.34
H461S	32 (11)	0.85 (0.03)	0.03	3.14 (0.54)	0.60 (0.04)	0.19
R520V	22 (3.6)	1.73 (0.09)	0.09	2.3 (0.5)	2.16 (0.03)	0.92

Standard deviation is in parenthesis.[a] an example of one batch, [b](Hibbert et al., 2007)

From Table 3.2, it can be seen that the previously determined specific activities of wild type and the single mutants were much lower than what have been identified in this study. The initial rates of all the reactions in the previous study were much lower although the enzyme concentrations used were much higher. The causes of the discrepancies may be the differences in the enzyme and substrate preparations. I also discovered that certain mutants were unstable while their expression levels fluctuated from batch to batch and depended on the culture condition (see later on). Propionaldehyde is also easily oxidised and degraded. Since more details in the activity assays for both substrates are not available in their publications (Hibbert et al., 2008, 2007), other causes of deviations cannot be determined. Therefore, the relative influences of each mutation here will be discussed based on comparison of the results from this study only. It should be noted that the specific activities of the single mutants in the publication Strafford*et al.*, 2012)differed from those listed in Table 3.1 because the specific activities in the publication were the averaged results from this work and all previous repeats (Hibbert et al., 2008, 2007)(Strafford, personal communication).

Prior to any discussion, two important observations must be noted as they have the potential to influence the apparent characteristic of the mutants. Firstly, the mutants in the first group strongly preferred glycolaldehyde over propionaldehyde as an acceptor substrate with an approximately 200-fold difference in the bioconversion rates. These differences made the reaction times of propionaldehyde much longer and it is possible that the mutants in the first group were subject to a higher degree of inactivation by propionaldehyde. This inactivation could also occur to the mutants in the second group to a lesser extent. This is because the specific activities of the mutants evolved for propionaldehyde were still 15-35 times slower than the activity of wild type towards

glycolaldehyde. This factor may be more pronounced for inherently more unstable mutants and lead to lower observed activities. However, this factor will be neglected throughout the study. Secondly, in this experiment, I found that the H461S mutation had a tendency to cause enzyme inactivation. When the H461S mutant was grown and prepared for the activity assay, half of the batches did not show any enzymatic activity. The specific activity also greatly fluctuated which resulted in a high standard deviation as showed in Table 3.1. In addition, occasionally inactivated enzymes were also observed in A29E/H461S, H461S/R520V, and A29E/H461S/R520V mutants suggesting that H461S may cause a global instability or time-dependent enzyme inactivation.

3.3.3. The locations of the targets residues, their functions, and interactions

Among all three target residues in group I, A29 was the only residue that is not part of the active site. It is actually located at the phosphate binding pocket of ThDP as showed in Figure 3.3. Therefore, amino acid substitution at this residue is unlikely to alter the substrate binding. However, replacing alanine with a negatively charged amino acid can cause electrostatic repulsion between the side chain and the phosphate group of the ThDP molecule which could affect the affinity between the enzyme and cofactor or the reposition of the ThDP. These could potentially result in lower enzyme activity as observed in this experiment.

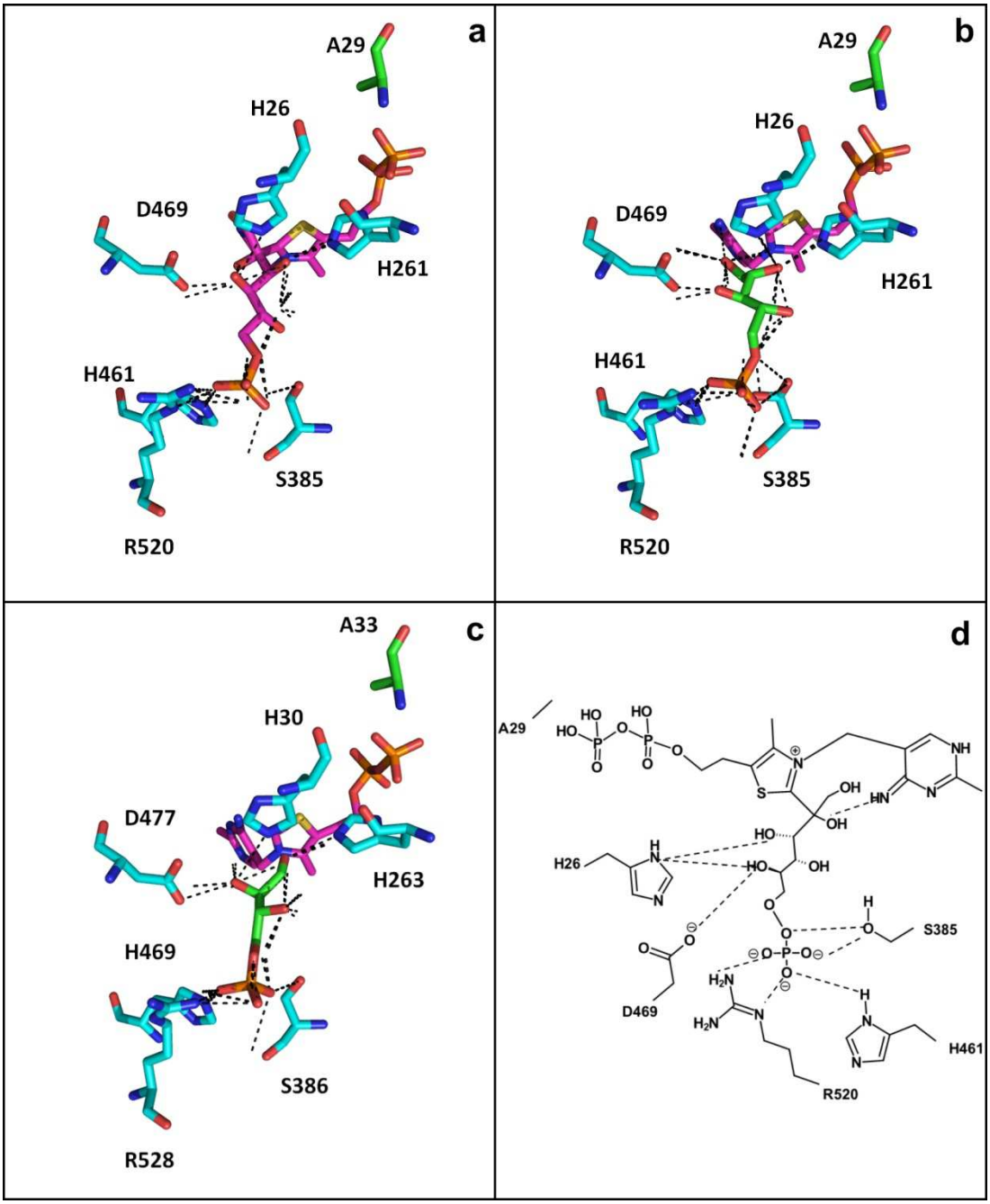


Figure 3.3 The location of the residues subject to mutation in this study. Black dotted lines are hydrogen bonds.

a) the covalent-linked intermediate between fructose 6-phosphate and ThDP in wild type TK from *E. coli*(PDB ID: 2R8P). b) the binding of ribose 5-phosphate into *E. coli* transketolase active site (PDB ID: 2R5N). c) the binding of erytrhose 4-phosphate into yeast transketolase active site (PDB IB: 1NGS). d) aschematic illustration of the hydrogen bonds between certain residues and covalent-linked intermediate between fructose 6-phosphate and ThDP.

The previous directed evolution study showed that replacing the positively charged phosphate-binding residues with neutral amino acids appear to improve the activity towards glycolaldehyde (Hibbert et al., 2007). In this study, however, the mutations at the

phosphate binding residues; H461S and R520V did not appear to improve the activity towards glycolaldehyde. In addition, all other single mutants in the first group were found to have lower activities towards propionaldehyde. Besides that, the single, double, and triple mutants constructed from combining H461S and R520V did not show a significant shift in substrate preference from hydroxylated to non-hydroxylated substrate. By analysis of the crystal structure of transketolase, it can be seen that these residues are located at the entrance of the transketolase active site and the distance from their amino groups to the C2 atom of the thiazolium ring are further than 10 Å. This distance is too far for the hydroxyl group of the glycolaldehyde to interact with while the carbonyl group is in close proximity to the C2 atom. This distance is, however, compatible with phosphorylated substrates as showed in the crystal structures of yeast and *E. coli* transketolases (Figure 3.3a and b) (Asztalos et al., 2007; Nilsson et al., 1997; Wikner et al., 1997). In the the crystal structure of *E. coli* transketolase (PDB ID: 2R5N, Figure 3.3b), it shows that the 3-hydroxyl group of ribose 5-phosphate forms a hydrogen bond with the side chain carboxylic group of D469 residue (Asztalos et al., 2007). On the other hand, D477 in yeast transketolase interacts with C2 hydroxyl group of a shorter acceptor molecule erythrose 4-phosphate (Figure 3.3c) (Nilsson et al., 1997). In addition, D469Y hada very low activity towards glycolaldehyde but significantly improved activity towards propionaldehyde, a dramatic change in the substrate preference from hydroxylated substrate to non-hydroxylated substrate as showed in Table 3.1. This combinded evidence suggest that the hydroxyl group of glycolaldehyde does not interact with the phosphate binding site. Therefore, it is possible that mutations at these residues may have small impacts on the activity towards glycolaldehyde or propionaldehyde.

Although none of the residues in groupI appeared to directly interact with glycolaldehyde, the mutation at H461 and R520 can cause some changes in the microenvironments in the active site channel or small structural arrangement which influences the overall activity and stability. In addition, there could also be synergy between these residues due to their spatially close positions. Therefore, the experimentally determined activities and expected activities were compared to find these interactions as showed in Figure 3.4a.

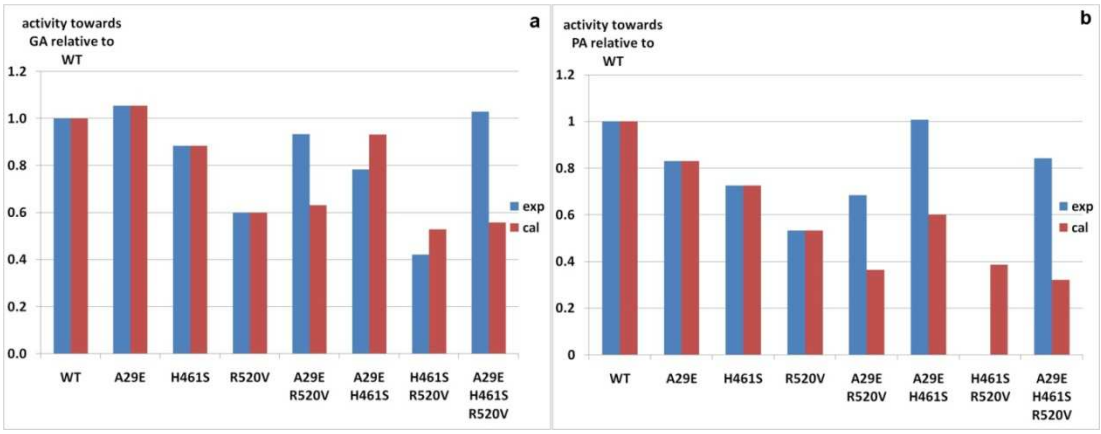


Figure 3.4 The relative activity of each mutant in group I towards a) glycolaldehyde and b) propionaldehyde. ■experimentally determined. ■calculated activities from additive synergy. All synergies were calculated from the method in Chapter 2, section 2.10using $\Delta G_{(I)} = RT \ln \{(A_{x,y} A_{WT})/(A_x A_y)\}$ equation.

From Figure 3.4a, the calculated activity of A29E/H461S and H461S/R520V showed that H461S exhibits negative synergy with A29E and R520V. However, the activity of the triple mutant is higher than the calculated one. This suggested that A29E may interact with R520V. Indeed, the experimentally determined activity of A29E/R520V and the triple mutant were higher than the calculated value suggesting positive synergy between the two residues. The same analysis was done with the activity towards propionaldehyde to see whether these behaviours only occur in glycolaldehyde. From Figure 3.4b, adding A29E to either H461S or R520V appears to have additive effect on the enzyme activity towards propionaldehyde. The interaction between A29E and R520V was similar to what has been observed in glycolaldehyde. The analysis of co-evolved network showed that A29 is not in any network (Strafford et al., 2012) and A29 itself is not conserved among transketolases (Hibbert et al., 2007). It is still possible that the effect of A29E mutation may propagate to the reactive centre through the repositioning of the ThDP molecule. Slight shift of the ThDP molecule, together with the reshaped active site, can therefore affect the apparent enzymatic activity. A28E/D469Y was also found to be inactive towards propionaldehyde which supported the hypothesis that both ThDP reposition and active site alteration can affect the enzymatic activity although both residues were not in the co-evolved network. In addition, A29E/D469Y can illustrate that any pair of mutant cannot be randomly combined.

While the H461S mutant has a low activity towards propionaldehyde, H461S/R520V totally lost its activity after storage at -80°C, although the activity towards glycolaldehyde could be determined prior to storage. This supports my observation that H461S brought a large degree of instability to the enzyme, and that this was sensitive to freezing and

thawing. The instability or loss in activity was not observed previously after mutagenesis of yeast transketolase at the equivalent residue (Nilsson et al., 1997). In the study of yeast transketolase, H469A still had 77% activity remaining, but the affinities towards phosphorylated substrates were much lower, which implied that this residue formed hydrogen bonds with the phosphate group of the substrate (Nilsson et al., 1997). However, different substitutions could also cause different effects on the enzyme function and stability. The function-instability relationship upon H461S introduction will be discussed with the expression study later on.

D469T/Y were previously identified with improved activity for propionaldehyde (Hibbert et al., 2008). The mutagenesis study in yeast transketolase together with the use of deoxy aldose sugars and an opposite stereoisomer at the C2, led to a proposal that Asp477 interacts with the hydroxyl group at the C2 of acceptor aldehyde which directly controls the stereoselectivity of the enzyme (Nilsson et al., 1998). In the *E. coli* TK structure (PDB ID: 1QGD) D469 and R520 do not form a direct hydrogen bond between them but rather via a water molecule and the backbone carbonyl group of L466. This also supports my speculation that H461 and R520 do not form any direct interaction with the 2-hydroxyl group of glycolaldehyde or the methyl group of propionaldehyde but the interactions would partly arise from the D469 residue (Hibbert et al., 2008). However, the effect of combining the mutants is still observable.

Some non-natural D469 variants, especially D469T, were found to have the highest activities towards propionaldehyde among several other mutants (Hibbert et al., 2008). D469T was previously combined with D259 mutants in order to further improve the activity. However, the double mutants D469T/D259S and D469T/D259Y had lower activities than the respective single mutants, while recombinants containing D469S did not suffer the same issue (Strafford et al., 2012). This led to a hypothesis of whether D469T was the cause of the loss of enzyme activities within the double mutants. In addition, a statistical coupling analysis of ThDP dependent enzyme sequences revealed that the D469 residue co-evolved with R520 (Strafford et al., 2012). Therefore, substituting R520 with other amino acids could potentially improve the enzyme activity or stability. Indeed this was observed, and in this study R520Q was chosen for further work because it was found to enhance the solubility but reduce the enzyme activity slightly (Strafford et al., 2012). In order to assess whether this was the case more generally, I recombined D469T/Y mutants with R520V/Q to study if each combination will have the same effects. None of these variants naturally exist

in transketolase which therefore has significant implications for enzyme engineering (Hibbert et al., 2007). In Figure 3.5, it can be seen that positive synergy was observed for D469T/R520Q whereas negative synergy was observed for D469Y/R520Q. However, when R520V was added into D469Y, only a slightly positive synergy was observed. This suggested that different substitutions at the first position (R520) can have a different effect on the second position.

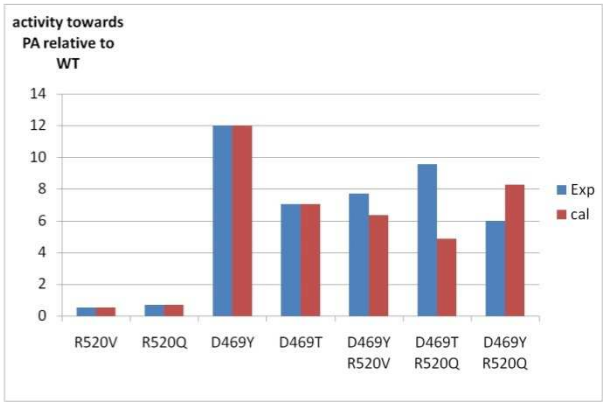


Figure 3.5 The relative activity of group 2 mutants towards propionaldehyde.

■experimentally determined ■ calculated activities from additive synergy.

3.3.4. H26Y, the inactive mutant

Another residue in the second group is H26Y. In the previous studies, H26Y was identified to reverse the stereoselectivity (Smith et al., 2008) and accept other aliphatic and cyclic aldehydes (Cázares et al., 2010).Other variants; A, T, K, and V were also found to have higher activity than wild type (Hibbert et al., 2008). In this study, however, H26Y was found to have no activity towards any substrate. Although D469Y can improve the activity towards propionaldehyde, the double mutant H26Y/D469Y did not restore the activity either. In addition, H26Y was reconstructed by another research group and was also found to have very little activity (Yi et al., 2012). In fact, H26 is highly conserved in transketolase-like enzyme family (Costelloe et al., 2008). The transketolase crystal structures from yeast (PDB ID: 1NGS) (Nilsson et al., 1997) and *E. coli* (PDB ID: 2R8O, 2R8P) (Asztalos et al., 2007) illustrated that H26 forms a hydrogen bond to the C3 hydroxyl groups of the donor molecules (Figure 3.3), which needs to be deprotonated during the catalytic cycle. Due to the close proximity and hydrogen bond interaction with the C3 hydroxyl group, H26 was proposed to abstract proton from this hydroxyl group (Wikner et al., 1997). The study of the equivalent histidine residue in yeast transketolase also showed that H30A/N mutants showed severely impaired k_{cat} (Wikner et al., 1997), although in the *E. coli* transketolase,

the H26A mutant was observed to have 2.3 times higher activity towards propionaldehyde than the wild type (Hibbert et al., 2008). Therefore, it is highly possible that H26Y will produce an inactive enzyme, particularly given the large steric bulk of the tyrosine residue.

The study in both groups of transketolase so far illustrated that combining high activity-mutants would not always improve the activity of the enzyme due to the synergistic interaction between these residues. Although the two distal residues are not in the same co-evolved network, they may appear to interact upon mutations but these interactions could possibly arise from a different reason to those driving the natural co-evolution (charge stabilisation, steric minimisation).

3.3.5. The expression level of all mutants

Enzyme activity is not the only parameter to be considered for synthetic purposes. Higher temperature or non-physiological pH can be encountered, and enzyme stability must also be considered. The effect of mutation does not only influence the enzymatic activity but also the stability of the enzyme. The *in vitro* stability of an enzyme can be determined from the thermodynamic stability study or kinetic stability study. In contrast, it was suggested that mutations within a protein can have an impact on protein folding mechanism which leads to the formation of inclusion bodies (Calloni et al., 2005). In the study of Strafford *et al.*, (2012), the expression levels of some single and multiple mutants were already assessed to investigate the instability upon the accumulated mutation(Strafford et al., 2012). Here, I further looked into the effect of the combination of the mutations in group I and II that I created to find out whether there was a relationship between their activities and stabilities. The expression level data were also employed in with the previous study to get a larger picture effect of the mutation on enzyme stability (Strafford et al., 2012). The mutation itself may promote the expression but not stability or folding. As a result, higher expression may not reflect higher stability. Therefore, I additionally looked into the ratio between the soluble and inclusion body content formed. The expression patterns of all the mutants were analysed and illustrated for comparison in Figure 3.6 below.

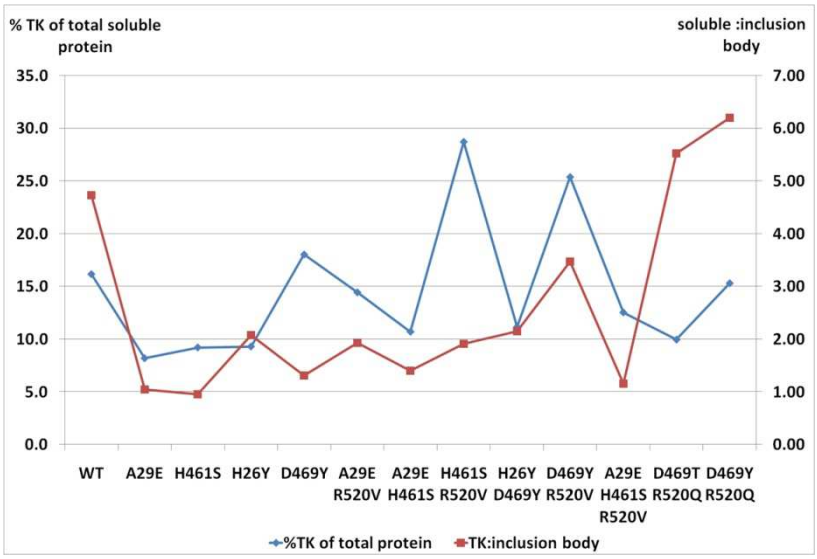


Figure 3.6 The fraction of transketolase.

◆ represents the percentage concentration of transketolase in the clarified lysate. ■ is the ratio of soluble TK : inclusion body.

In order to avoid a lower expression due to plasmid instability, the percentage expression was determined with the master glycerol stocks as these were not obtained from sub-cultures but rather represent the first generation after transformation in which the plasmid copy number is expected to be at its highest. Although the expression of all the mutants appeared to be at least 8 %, some quickly lost their expression levels even though the same glycerol stock was repeatedly used. These include H26Y, H26Y/D469Y, H461S, and H461S containing mutants. For H26Y and H26Y/D469Y, no enzyme was produced after the second time that the same glycerol stock was used. Since these phenomena only occurred in certain mutants, it suggested the rapid reductions in the expression level of these mutants were not due to the overall instability of the glycerol stock but rather the instability or the toxicity of the mutants themselves.

The blue line in Figure 3.6 illustrates that most mutants had lower expression levels than the wild type, while the soluble: inclusion body ratio (red line) decreased below that of the wild type. This suggested that larger fractions of the produced enzymes were in the inclusion body and implied that these mutants are beginning to become unstable. However, some double mutants appear to restore the solubility of the enzyme.

As mentioned above, H461S containing mutants quickly lost their expression levels when the same glycerol stock was repeatedly used. It is not surprising that several single mutations cause significant reduction in both expression level and activity. It has been

observed that a single mutation could change several properties of a protein simultaneously (Lee and Goodey, 2011). The only reported role of H461 is to bind to the phosphate group of the natural substrates (Asztalos et al., 2007; Nilsson et al., 1997; Wikner et al., 1997). In order to rationalise the cause of enzyme in activation upon H461S mutation, the crystal structure analysis was employed. However, the crystal structures of both *E. coli* and yeast transketolases show that the side chain of H461 did not form extensive hydrogen bond with other nearby residues. Consequently, the explanation why H461S would destabilise or inactivate the enzyme may not be sought from structural information. In fact, the structural information alone cannot provide the explanation of all stabilisations due to the subtle changes on the structure upon mutation. It was found through the evolutionary analysis that H461 is part of the co-evolved network between PP-Pyr domain (Strafford et al., 2012), so it is possible that this network was disrupted. It has been suggested that if the coupling residue is also mutated, the function can be rescued (Neher, 1994). In this study, R520 was the only coupling residue of H461 and only R520V was combined with H461S while other possible sites and variants were not explored. In order to confirm this possibility, other residues have to be explored.

Among several low expression mutants, the expression level of R520V mutant was very sensitive to the culture conditions. When this mutant was cultured in a falcon tube under the same culture conditions as other mutants, very low expression (<5%) was observed and a significant amount of inclusion body was detected. However, when switching the fermentation condition to 250-ml shake flask with 20-ml culture volume, the expression was recovered to above 15%. Although the single mutant R520V seemed to cause low expression level, double mutants bearing R520V appeared to have expression levels similar to or above the wild type. In A29E/R520V, not only the expression level was improved but also the soluble: inclusion body ratio. Although the same effects seemed to happen in H461S/R520V, this mutant quickly lost its expression level which implied that R520V cannot compensate the instability brought by H461S. Besides that, A29 was neither part of the highly conserved residue (Hibbert et al., 2007) nor within the co-evolved network (Strafford et al., 2012) which may explain why it can overall tolerate amino acid substitution better than H461 and R520.

The interaction between D469 and R520 upon mutation was also observed in the expression level study. R520V also shows to enhance the solubility of D469Y/R520V as more soluble transketolase and higher soluble : inclusion body ratio than D469Y were

observed. However, at least 20 % of the total D469Y/R520V produced by the cell was in the form of inclusion body. R502Q, in contrast, significantly reduced the amount of inclusion body in D469T/R520Q and D469Y/R520Q to less than 15% which increased the soluble : inclusion body ratio above the wild type. However, the % of total protein for D469Y/R520Q in the lysate is much lower than D469Y/R520V. It could be said that R520V enhances the production of the mutant within the cell whereas R520Q slightly reduces the total production but facility the folding and the formation of soluble enzyme. This is another difference between R520V and R520Q mutation on D469Y.

H26Y and H26Y/D469Y although showed 9.2 and 11 % expression level of the total soluble protein in Figure 3.6, their expression can only be detected straight after transformation. The expression level in the same glycerol stock that has been used for streaking twice declined to less than 5%. I initially suspected that H26Y and H26Y/D469Y were not stable and may precipitate upon sonication or form a large amount of inclusion body during cell culture. Therefore, both mutants were cultured for 24 hours at 25°C with and without 10% ethanol supplement in the media. This is to slow down cell growth and allow more time for the protein to fold. Ethanol was added to induce heat-shock protein to aid protein folding (Thomas and Baneyx, 1997, 1996). The half of the harvested cells was then lysed by freeze-thaw lysis while the other half was sonicated. None of these have improved the expression level of both mutants. The SDS-PAGE of the cell debris did not show that these mutants form any inclusion body either which suggested that both H26Y and H26Y/D469Y did not express. These evidences imply that H26 is not only important for enzymatic activity but may also the control of transketolase expression.

The loss of expression level and activities in certain mutants may suggest that these mutants may have lost higher order structure or become soluble multimeric enzymes. In order to see whether this is the actual cause, these mutants could be purified and the distribution of the molecular weights analysed by gel filtration. It should still be noted that the purification process itself may inactive transketolase or cause it to aggregate as shown by Stratford et al. (2012).

During the study of the insoluble fraction of all the mutants, one band always appeared in the cell debris samples but showed differences in the expression level. The size of this protein was determined from the log(molecular weight) versus the migration distance and was estimated to be approximately 40 kDa. In order to identify whether this is a fraction transketolase, the sample was sent for protein sequencing. The result showed

that the first 5 amino acids from the N terminus are AEIYN which are not found in transketolase sequence. This sequence was then BLAST search in BL21(DE3) (taxid:469008). Table 3.3 shows the proteins with AEIYN sequence and the position of this sequence.

Table 3.3 The possible P2 proteins.

Protein AC/ID	Name	Length (aa)	Match range
C6EA44/C6EA44_ECOBD	Outer membrane phosphoporin protein E precursor	345	22-26
C6EER5/C6EER5_ECOBD	GCN5-related N-acetyltransferase precursor	172	15-19
C6EI53/C6EI53_ECOBD	OmpF, subunit of outer membrane porin F and The Colicin A Import System precursor	362	23-27
C6EKF5/C6EKF5_ECOBD	Flavohemoprotein	396	136-140
E5QQ21/E5QQ21_ECOBD	Outer membrane porin protein, N-ter	201	24-28

From the position of the sequence and the molecular weight of the protein, the most likely protein is OmpF. OmpF is associated with stress response and control the material across the cell membrane. In a low nutrient environment, high expression level of OmpF is to be expected to facilitate nutrient uptake (Delihas and Forst, 2001). Therefore, it could be argued that OmpF may be highly expressed towards the end of the culture when the nutrient was low. Therefore, P2 concentration was plotted against the harvest OD600. However, this plot has no strong correlation to the expression level of OmpF. The high expression is unlikely to be due to low nutrient level at the end of culture. I further investigated to see whether the expression of this protein is related to the formation of the transketolase inclusion body by plotting the P2 concentration against soluble: inclusion body ratio as showed in Figure 3.7. The P2 concentration has an inverse relationship with the soluble: inclusion body ratio. The higher ratio arose from either low soluble fraction or high inclusion body and both situations can trigger stress response within the cell.

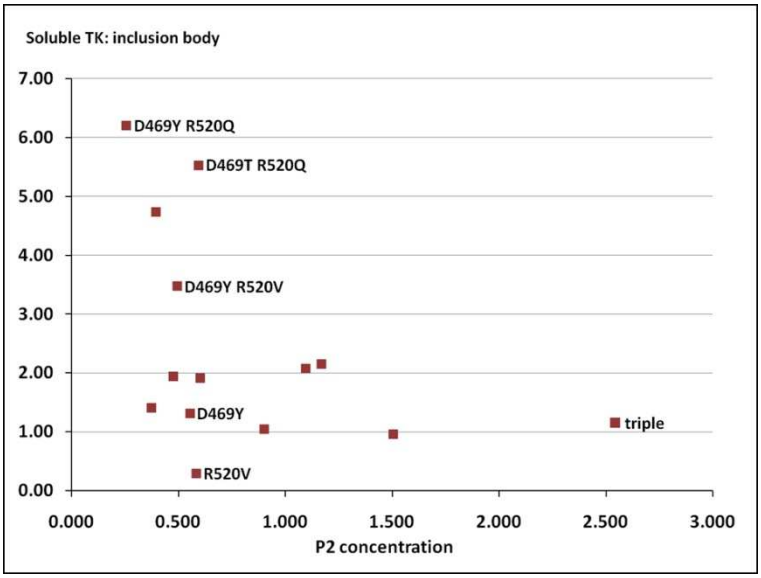


Figure 3.7 The scatter plot between the P2 concentration and the TK soluble: inclusion body ratio.

3.3.6. Correlation of expression and activity

As part of the aim of the study in the publication (Strafford et al., 2012), the activity and stability relationship of the mutants was investigated and a trade off between the enzyme activity and stability was observed. All the expression and activity data here were compiled to compare this relationship when the mutants were constructed based on two different approaches. All the mutants in this study were created from combining the single mutants which were previously reported to be highly active towards glycolaldehyde (Hibbert et al., 2007) and propionaldehyde (Hibbert et al., 2008). In the previous study, some double mutants were the created with the same approach while some double mutants were constructed by choosing target residues within the co-evolved network (Strafford et al., 2012). Assuming that the % TK concentration in the clarified lysate represents the stability of the protein, I plotted the % TK concentration versus activity towards propionaldehyde to see if this trend could be observed as showed in Figure 3.8a. Since group I mutants strongly preferred glycolaldehyde over propionaldehyde, their expression levels were also plotted against their activities towards glycolaldehyde (Figure 3.8b).

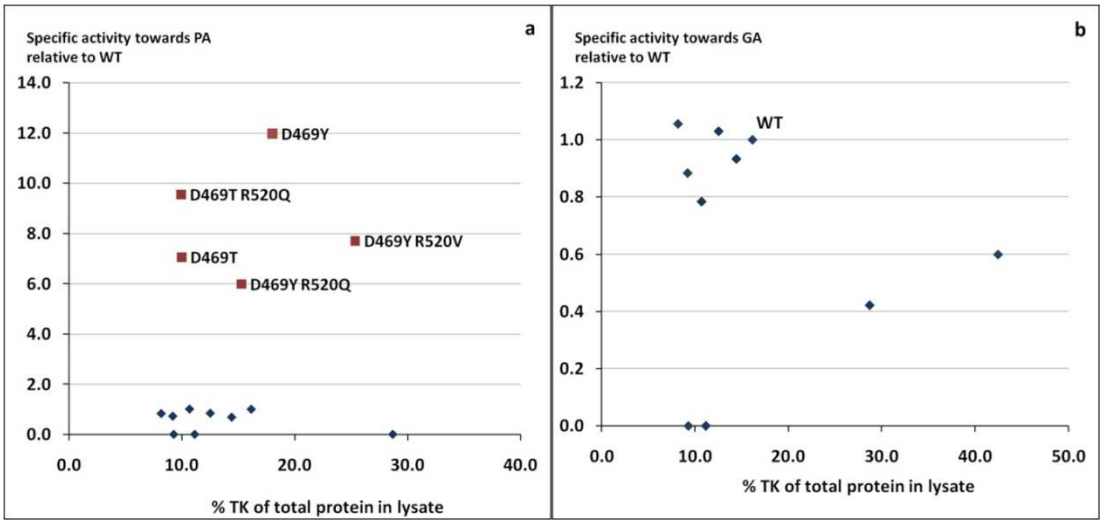


Figure 3.8 The relationship between the specific activity towards glycolaldehyde and propionaldehyde and the expression level of the mutant.

◆mutants in group 1. ■mutants in group 2

Unfortunately, the number of mutants from this study alone is insufficient to draw this conclusion. In addition, several mutants in this study contain a global inactivating mutant; H461S, and catalytically important residue; H26, which can disrupt the relationship between stability-activity. Furthermore, drastic decreases in expression levels of several mutants were observed within a few times that the same glycerol stock was used. Including such uncertain-characteristic mutants in the analysis would lead to inaccurate interpretation of the activity-stability relationship.

3.4. Conclusion

Here we have illustrated that most mutants that were created by pairwise recombining two previously active single mutants within the enzyme active site resulted in a rapid loss of expression level, suggesting that this recombination strategy could have a detrimental effect on the enzyme stability. There could also be strong synergy between them although this was not always showed by analysis of the natural co-evolved network. Since the consistent expression level was only observed in the double mutants in which both residues were within the same co-evolved cluster, this method of designing a library could be more attractive. In addition, a single mutation may have negative effect on the enzyme but this could have a positive effect when recombining with other residues in the coupled network.

In this study, we also found that the single mutant H461S appears to globally destabilise or inactivate the enzyme. This destabilisation cannot be deduced from the crystal structure information but the intrinsic stabilising function may only be determined from mutagenesis and statistical analysis of kinetic data, which has to be further studied.

4. Chapter 4: Rational design of substrate and TK for aromatic aldehyde

4.1. Introduction

Dihydroxy ketone functionalilty is found in a number of high value, natural compounds (Enders and Narine, 2008; Enders et al., 2005; Wohlgemuth, 2009) such as carbohydrate (Kobori et al., 1992), and cortocosteriods (Hailes et al., 2009; Hailes et al., 2010). The stereoselective synthesis of these chemicals, however, is rather complicated, low atom efficient, and involves protecting groups and requires multi-steps (Hailes et al., 2009; Hailes et al., 2010). Transketolase catalyse the asymmetric carbon elongation by transferring dihydroxy ketone group to an aldehyde to form dihydroxy ketone product with high stereospecicity and selectivity and this makes transketolase attractive in organic synthesis.

The natural substrate spectrum of wild type transketolase from several species includes non-phosphorylated sugar (Demuynck et al., 1991; Kobori et al., 1992), non-hydroxylated aldehydes, cyclic aldehydes (Hobbs et al., 1993; Morris et al., 1996), and nitroso aromatics (Corbett and Corbett, 1986). This makes the earlier work with transketolase focus on the synthesis of ketose sugars (Bolte et al., 1987), deoxysugars (Hecquet et al., 1994) which could be a building block for other compounds when integrated with other enzymatic systems or chemical routes. These include the synthesis of a caramel-like flavour; furaneol, (Hecquet et al., 1996; Hecquet et al., 1994), glycosidase inhibitors; 1,4-Dideoxy-1,4-imino-D-arabinitol (Ziegler et al., 1988) and *N*-hydroxypyrrolidine (Humphrey et al., 2000), beetle pheromone (+)-*exo*-brevicomin(Myles et al., 1991), and 2-amino 1,3,4-butanetriol which is a building block for an antiviral drug; Nelfinavir (Ingram et al., 2007).

Transketolase could be coupled with transaminase where the dihydroxy ketone is aminated by transaminase to produce chiral amino alcohols such as glycosidase inhibitor, sphingolipids, and chlorampenicol and analogues(Hailes et al., 2010; Hailes et al., 2010; Smith et al., 2010). In the case of chloramphenicol synthesis, the reaction has to start with benzaldehyde derivatives (Hailes et al., 2010) which is not accepted by wild type *E. coli* transketolase (Galman et al., 2010) and very low activities were reported with wild type spinach transketolase (Demuynck et al., 1991).

Transketolases from many species prefer hydroxylated over non-hydroxylated substrates (Demuynck et al., 1991; Hobbs et al., 1993; Morris et al., 1996) due to the extensive hydrogen bonds with conserved histidine and aspartate residues within the active site channel (Asztalos et al., 2007; Wikner et al., 1997) and these network makes transketolase more readily for engineering towards these substrates. Among these residues, D469 was randomly substituted and several variants show significant enhanced activity towards propionaldehyde (Hibbert et al., 2008) with higher enantiomeric excess (Smith et al., 2008). Several variants were also reported to accept longer chain aldehydes with high stereoselectivity (Cázares et al., 2010). Recently, D469T, D469K, and D469E were found to be able to catalyse aromatic and heteroaromatic aldehydes with the yield below 10% and side reactions (Galman et al., 2010). The low yields were suggested to be due to steric hindrance upon substrate binding which lead to the rational design of F434A mutant which shows to improve the overnight yield of benzaldehyde to 10% with no side reaction (Galman et al., 2010).

This chapter aims to identify the causes of low bioconversion of aromatic aldehydes by transketolase. Benzaldehyde does not have any functional group that could form a specific interaction with the active site of transketolase. Therefore, 3-formylbenzoic acid and 4-formylbenzoic acid were chosen in order to reconstitute the hydrogen bonds between the highly conserved phosphate binding sites and the substrates as observed in natural substrates. Both substrates were screened against all the mutants that were reported to accept benzaldehyde and 3-hydroxybenzaldehyde to investigate the residues that contribute the steric hindrance. The mutant that accepts these substrates was then subject to mutagenesis at the phosphate binding sites and kinetic study in order to identify their influences in the bioconversion of aromatic substrates. The experimental results were compared with computational docking of the both substrates in the modelled mutant structures.

4.2. Materials and methods

4.2.1. Chemicals and reagents

All the reagents were purchased from Sigma Aldrich, UK, unless otherwise stated.

4.2.2. Transketolase library

Transketolase mutants were expressed under the same expression system as the wild type of which the details were provided in Chapter 2. The sequences of the forward primers are listed below. The target codons are in bold letter and the substituted bases are underlined.

R358P: GAAAATCGCCAGC**CCA**AAAGCGTCTCAGAATG

R358L: GAAAATCGCCAGC**CTT**AAAGCGTCTCAGAATG

H461S: GGTGATGGTTTACACC**AGC**GACTCCATCGGTCTGG

D469Y: TCGGTCTGGGCGAA**TAC**GGGCCGACTCACCAG

R520Q: GATCCTCTCCCAG**CAG**AACCTGGCGCAG

F434A: CGTACACCTCCACC**GCC**CTGATGTTCGTGG

D469T: TCGGTCTGGGCGAA**ACC**GGGCCGACTCACCAG

The plasmids of all the mutants were extracted and sequenced to confirm the amino acid substitution according to the standard protocol.

4.2.3. Enzyme preparation and quantification

Transketolase mutants were cultured according to the standard protocol in chapter 2. The fermentation volume of all the cultures was 50 ml which was performed in 250-ml flask. The total protein concentration was quantified by Bradford assay and the transketolase concentration was identified by densitometry.

4.2.4. Screening assay when using 3-FBA and 4-FBA as a substrates

Wild type, the mutants with enhanced activities towards non-hydroxylated substrates; D469Y, D469T (Cázares et al., 2010; Galman et al., 2010; Hibbert et al., 2008; Smith et al., 2008; Strafford et al., 2012) and the mutant that accepts aromatic substrates; F434A (Galman et al., 2010) were screened for 3-FBA and 4-FBA. 30 µl of clarified lysate was incubated with 30 µl of 10x-cofactor solution (24 mM ThDP, 90 mM $MgCl_2$ in 50 mM Tris buffer, pH 7.0) and 140 µl of 50 mM Tris buffer, pH 7.0 for 20 minutes in 1.5 ml glass vials. The reaction was started by the addition of 100 µl of 3x substrate solution (150 mM HPA and 150 mM 3-FBA or 90 mM HPA and 90 mM 4-FBA in 50 mM Tris buffer, pH 7.0) which was prepared in glass vial. 24 hours after the reaction starts at 25 °C, 20 µl of the reaction sample was quenched in 180 µl of 0.1% TFA. The samples were centrifuged at

13,000 rpm for 3 minutes. The supernatants were analysed by HPLC with Aminex HPx-87H column and UV detection at 210 nm. This is the same system as for the detection of erythrulose and HPA but the retention time was extended to 120 minutes. The retention times of 3-FBA and 4-FBA were 64 minutes.

4.2.5. Enzyme kinetics for 3-FBA and 4-FBA

All the reactions were performed in glass vials at 25 °C in triplicate. Due to the instability of these mutants, clarified lysate was used throughout the study instead of the affinity purified enzyme. 3x HPA and aldehyde substrate stock solutions were freshly prepared prior the reaction. All substrates were prepared in 50 mM Tris buffer and adjusted the pH to 7.0. The concentration of HPA was maintained constant at 50 mM in all the reactions. Transketolase concentrations within the reactions were between 0.04 – 0.07 mg/ml. Due to the inhibition above 50 mM, the concentrations of 3-FBA in the reaction were varied between 10 – 40 mM. The concentrations of 4-FBA in the reaction were between 6 – 30 mM. All the enzymes were prepared in the same way as for the screening. The triplicate reactions were started by adding 100 µl of the 3x substrate stock solution at 15-second interval. After adding the substrates, the vials were quickly closed by snap caps. 20 µl samples of 3-FBA were taken every 3 minutes for 15 minutes and quenched in 380 µl 0.1% TFA. The samples were centrifuged and the supernatants were analysed by HPLC with ACE5 C18 reverse phase column (150 × 4.6 mm). The details of the mobile phase are in Chapter 2, section 2.7.2. The retention times of 3-FBA and its product are 5.47 and 2.97 minutes, respectively. Michaelis-Menten plot was used to determine the K_M and k_{cat} of 3-FBA. Instead of plotting the initial rate against the substrate concentration, specific activities at each concentration were plotted in order to combine the repeated data. All the data were plotted into SigmaPlot 12.0 as illustrated in Chapter 2, section 2.9. The unit of µmol mg^{-1} min^{-1} was converted to s^{-1} as below in order to calculate the k_{cat} as showed in Equation 4.1

$$\mu\text{mol mg}^{-1}\ \text{min}^{-1} = 10^{-6}\ \text{mol} \times \left(\frac{72197\ \text{g mol}^{-1}}{10^{-3}\ \text{g}}\right) \times \frac{1\ \text{s}^{-1}}{60}$$

$$\text{specific activity } (\text{s}^{-1}) = \frac{k_{cat}[S]}{K_M + [S]}$$

Equation 4.1 The modified Michaelis-Menten equation used for the calculation of the kinetic parameters

The 4-FBA reactions were performed in a similar manner but the samples were taken every 30 minutes. Due to a slower reaction, the enzyme concentrations within the reaction were 0.104 mg/ml for D469T and 0.171 mg/ml for D469T/R520Q. The retention times of 4-FBA and its product are 5.41 and 2.45 minutes respectively. The K_M and k_{cat} of 4-FBA were determined in the same manner. However, due to the linearity of the data when using D469T, a double reciprocal Lineweaver Burk was also used to compare the kinetic parameters.

Standard substrates and the dihydroxy-ketone products of 3-FBA and 4-FBA were used to quantify the conversion yields at different time points.

4.2.5.1. Standard graphs of 3-FBA and 4-FBA

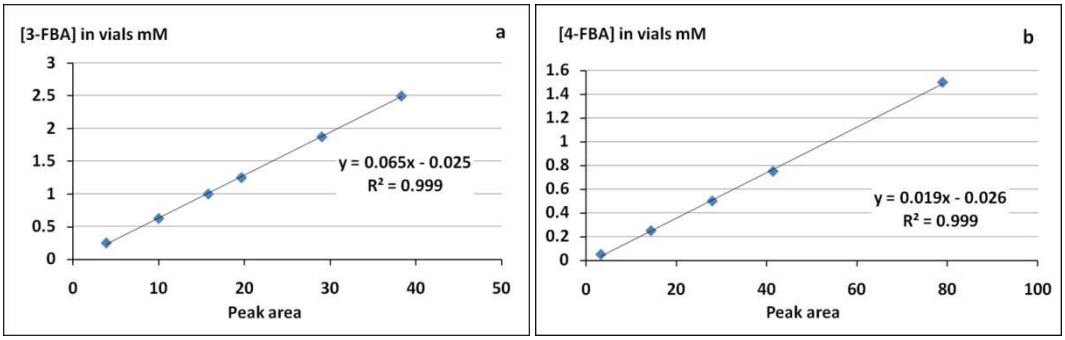


Figure 4.1 The standard graph of 3-FBA and 4-FBA.

The standard solutions were analysed by ACE5 C18 column and monitored by UV absorption at 275 nm. The graphs were plotted basedon the peak areas and the final concentration of the aldehydes in the vials.

4.2.6. Propionaldehyde activity

Specific activity towards propionaldehyde was determined using clarified lysate which was prepared in the same way as the screening method. The specific activities were assessed at 50 mM HPA and 50 mM propionaldehyde. The activity assay was performed according to the standard protocol in chapter 2.

4.2.7. Synthesis of 3- and 4-(1,3-Dihydroxy-2-oxopropyl)benzoic acid

Both are the dihydroxy ketone products of 3-FBA and 4-FBA. They were prepared by David Steadman (Chemistry Department). The details of the preparation and characterisation are available from the publication (Payongsri et al., 2012).

4.2.8. Computational docking of 3-FBA and 4-FBAinto TK mutant structures

The modelled D469T and D469T/R520Q structures were prepared by introducing mutations into the *E. coli* transketolase crystal structure (PDB ID: 1QGD). Residues at 10 Å around the ThDP molecule were selected and then energy minimised using Charm forcefield Adopted Basic NR, Implicit Generalised Born solvent model, True SHAK E constant, and 1000 steps in Discovery Studio 2.0 (Accelrys, Inc. San Diego, California, USA). The energy minimised ThDP-enamine structure and 3-FBA or 4-FBA were then docked into these energy-minimised structures using AutoDock 4.2 and Auto-Dock tools 1.5.4. This computational docking was performed by David Steadman (Chemistry Department).

The binding energies of 3-FBA and 4-FBA were used to calculate the K_d values of the substrates and compared with the K_M values of all the acceptor substrates that have been previously reported. All the calculated K_M values of other natural substrates were performed by John Strafford (former PhD at Biochemical Engineering Department).

4.2.9. Reaction of D469T with 3-nitrobenzaldehyde (unpublished data)

The nitro group on 3-nitrobenzaldehyde is a less electronegative group than the carboxylic acid group on 3-FBA and 4-FBA. This substrate was used to study how the strength of the negative charge on the substrate influences the enzyme activity. Only 3-nitrobenzaldehyde consumption was monitored in this study.

Due to the low solubility of 3-nitrobenzaldehyde, 20 mM substrate solution was prepared and used as 2x stock solution. 30.2 mg of 3-nitrobenzaldehyde was added into 10 ml of 50 mM Tris buffer. The mixture was stirred in a 60 °C water bath until all the powder dissolved. The solution was then left to cool down to room temperature. This solution was then used to dissolve HPA to the final concentration of 100 mM and the pH of this 2x stock solution was adjusted to 7.0.

30 µl of clarified lysate was incubated with 30 µl 10x cofactor solution (as above) and 90 µl 50 mM Tris buffer, pH 7.0 for 20 minutes. 150 µl of the 2X substrate solution was added to the enzyme-cofactor solution to start the reaction.

To monitor the reaction, 20 µl samples were taken at time zero, 5 hours, 24 hours, and 48 hours and added into 380 µl of 0.1% TFA. The samples were centrifuged and the supernatants were analysed by HPLC with ACE5 C18 reverse phase column (150 × 4.6 mm). The flow rate was maintained at 1.0 ml/min. The mobile phase uses 2 solvents which are

0.2 M acetic acid and 80% (V/V) methanol. The flow profile comprises of 3 phases. In the 1st phase, the mobile phase was maintained at 90% acetic acid and 10 % methanol (80% V/V) for 5 minutes. In the 2nd phase, the composition was linearly changed to 40% 0.2 M acetic acid and 60% methanol over 14 minutes. This final point was maintained for 5 minutes in the 3rd phase. This was then followed by a 3-minute equilibration step which consists of 90% 0.2 M acetic acid and 10% methanol at 1.0 ml/min. 3-nitrobenzaldehyde substrate was eluted at 17.5 minutes.

4.3. Results and discussions

4.3.1. Screening for 3-FBA and 4-FBA

Transketolase crystal structures and previous mutagenesis studies revealed that there are several highly conserved residues that are essential for substrate specificity including phosphate binding residues (Nilsson et al. 1997; Schneider and Lindqvist 1998; Asztalos et al. 2007). Among several natural substrates, ribose 5-phosphate and fructose 6-phosphate naturally exit in cyclic structures. In addition, the cyclic structure of ribose 5-phosphate was also found in the active site of the crystal structure of *E. coli* transketolase (PDB ID: 2R5N) as showed in Figure 4.2a. This suggests that previously generated transketolase mutants may be able accept the aromatic six-membered ring of benzaldehyde. However, the lack of specific interaction in benzaldehyde can lead to low enzyme-substrate affinity, hence, low activity. This was highlighted when comparing the K_M values between the phosphorylated and non-phosphorylated substrates where the K_M values of non-phosphorylated substrates are at least 3 orders of magnitude higher than the phosphorylated ones whereas their k_{cat} values are similar as showed in (Table 4.1) (Nilsson et al., 1998; Sprenger et al., 1995). I hypothesised that the negatively charged carboxylic acid at the 3- or 4- position on the benzaldehyde ring could enhance the enzyme-substrate affinity. This was later supported when comparing the docking of 3-FBA and 4-FBA in the D469T and D469T/R520Q modelled structures with the *E. coli* transketolase structure with bound cyclic ribose 5-phosphate (PDB ID: 2R5N).

Table 4.1 The K_M and k_{cat} values of yeast and *E. coli* transketolases towards different acceptor substrates.

TK	Acceptor substrate	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M s^{-1} M^{-1}	Reference
E. coli TK	Ribose-5P	1.4			(Sprenger et al., 1995)
Yeast TK	Ribose-5P	0.146	46.3	$3.2x10^5$	(Nilsson et al., 1997)
E. coli TK	Erythrose-4P	0.090			(Sprenger et al., 1995)
Yeast TK	Erythrose-4P	0.0406	69	$1.7x10^6$	(Nilsson et al., 1998)
E. coli TK	Glycolaldehyde	14			(Sprenger et al., 1995)
E. coli TK	Glycolaldehyde	35	9.9	283	(Hibbert et al., 2007)
E. coli TK	Ribose	1400			(Sprenger et al., 1995)
E. coli TK	Erythrose	150			(Sprenger et al., 1995)
E. coli TK D469T	Propionaldehyde	55	5	91	(Hibbert et al., 2008)

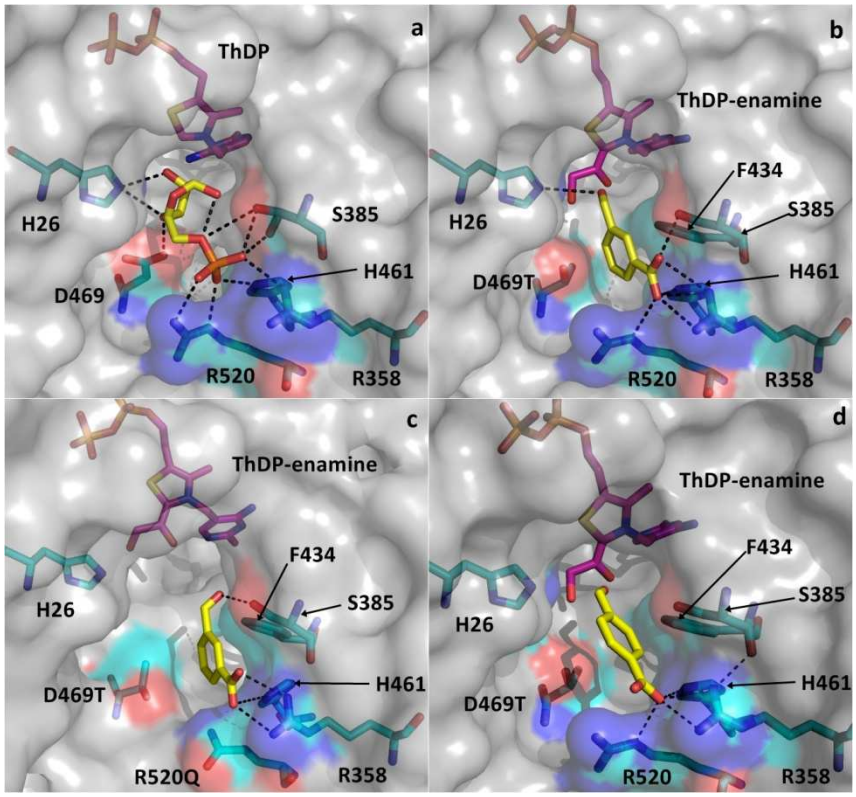


Figure 4.2 The ribose 5-phosphate-bound *E. coli* transketolase crystal structure and the computational docking of 3-FBA and 4-FBA into the energy-minimised transketolase mutant structures.

a) The crystal structure of wild type *E. coli* transketolase with cyclic ribose 5-phosphate (PDB ID: 2R5N). b) The computational docking of 3-FBA into D469T. c) The computational docking of 3-FBA into D469T/R520Q. d) The computational docking of 4-FBA into D469T.

The crystal structure of *E. coli* transketolase with cyclic ribose 5-phosphate showed that the phosphate group of ribose 5-phosphate interacts with R358, S385, H461, and R520 (Asztalos et al., 2007) and these interactions were also observed in yeast transketolase in complex with erythrose 4-phosphate (Nilsson et al., 1997). Computational docking of 3-FBA into D469T also suggested that the carboxylic acid group also interact with those phosphate groups in a similar manner. In addition, the orientation of the aromatic ring is also similar to the ring of ribose 5-phosphate. With the pK_a values of 3.8 and 3.7 for 3-FBA and 4-FBA,respectively (Jover et al., 2008),both substrates are fully ionised and bear negatively charged carboxylate group at pH 7.0 which is the reaction condition. Therefore, both substrates can potentially restore the electrostatic interaction with the phosphate binding residues at the active site.

In order to investigate whether 3-FBA and 4-FBA bind to the active site in the predicted manner and to see if other mutants would accept 3-FBA and 4-FBA, both substrates were tested with wild type and previously constructed mutants; D469Y, D469T,

and F434A and compare their bioconversions with benzaldehyde and 3-hydroxybenzaldehyde as showed in Table 4.2. After 24 hours, only D469T showed significant reductions in 3-FBA and 4-FBA concentration with 65% and 30% conversion yields, respectively. The chromatograms of the samples from both reactions were showed in Figure 4.3 a-d. The conversion yields were much higher than any previously reported yield with benzaldehyde derivatives (Table 4.2). The time course sampling revealed that the bioconversion of 3-FBA by D469T reached 65% conversion after 2 hours whereas 2% conversion of benzaldehyde was achieved after 17 hours. This suggests that the presence of the carboxylic acid on 3-FBA could restore the hydrogen bond and electrostatic interactions with the enzyme which had sped up the reaction of 3-FBA by D469T to 250 times faster than the reaction of benzaldehyde. It should be noted that the slow reaction could also lead to enzyme inactivation by the left over aldehyde which could further reduce the conversion yield.

Table 4.2 The bioconversion of different aromatic aldehydes by transketolase mutants.

Mutant	Reported isolated yield[a]		Conversion yield[b]	
	BA	3-HBA	3-FBA	4-FBA
WT	0%	0%	0%	0%
D469T	2%	4%	65%	30%
D469Y	4%	n.d.	0%	0%
D469K	2%	n.d.	n.d.	n.d.
D469E	2%	0%	n.d.	n.d.
F434A	10%	6%	0%	0%

n.d. is not determined.[a](Galman et al., 2010).[b] This study.

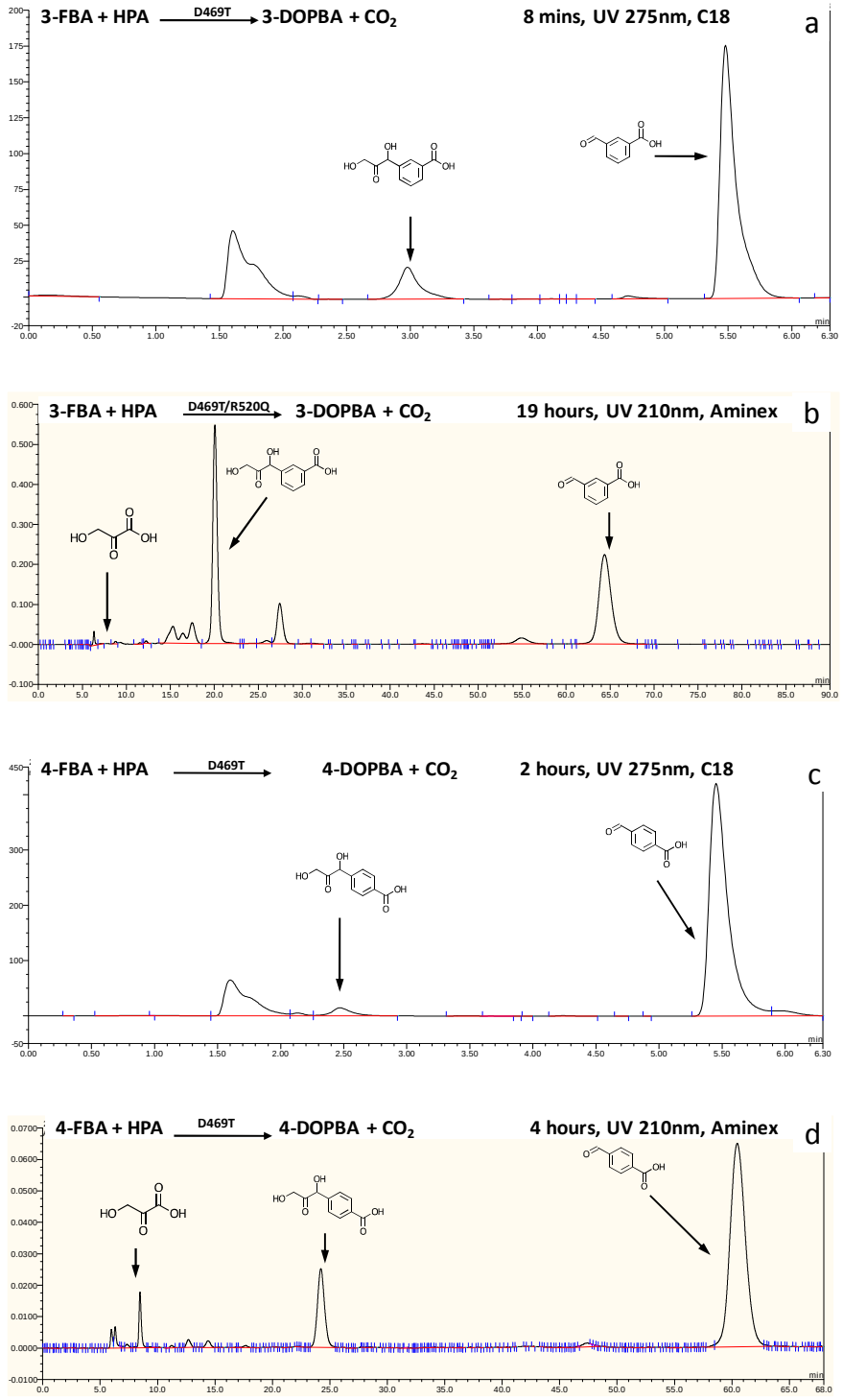


Figure 4.3 The Chromatograms of 3-FBA and 4-FBA reactions when analysed by ACE5 C18 reverse phase column (150 × 4.6 mm) and Aminex HPx-87H column respectively.

a) The 8-minute sample from 3-FBA reaction catalysed by D469T, analysed by ACE5 C18 column. b) The 18-hour sample from 3-FBA reaction catalysed by D469T/R520Q, analysed by Aminex column. c) The 2-hour sample from 4-FBA reaction catalysed by D469T, analysed by ACE5 C18 column. d) The 4-hour sample from 4-FBA reaction catalysed by D469T, analysed by Aminex column. All reactions were started with 50 mM of HPA and 50 mM aldehydes.

Several D469 mutants are capable to catalyse benzaldehyde and reach similar yields but only D469T was reported to accept 3-hydroxybenzaldehyde(Galman et al., 2010). Crystal structures of yeast and *E. coli* transketolase reveal that D469 residue interacts with hydroxyl group of the substrates (Nilsson et al., 1998; Smith et al., 2008). Replacing this negatively charged amino acid with a more hydrophobic amino acids such as threonine and tyrosine was reported to enhance the activity toward non-hydroxylated substrates due to the more favourable interactions with non-polar groups on substrates (Cázares et al., 2010; Hibbert et al., 2008; Smith et al., 2008). This favourable hydrophobic interaction may be one of the key factors that allow D469T to catalyse aromatic aldehydes that were not accepted by wild type. In addition, the side chain of threonine is also shorter than the side chain of aspartate and replacing aspartate with threonine can widen the active site and relieve the steric hindrance. This is particular true if the aromatic ring aligned in the active site as predicted in Figure 4.2 where D469 is at the bottom of the ring.

Although D469Y was one of the mutants with the highest activity for propionaldehyde by providing a similar hydrophobic interaction (Hibbert et al., 2008; Smith et al., 2008; Strafford et al., 2012), it was unable to catalyse the bioconversion of neither 3-FBA nor 4-FBA. By considering the size of the side chain, it can be seen that tyrosine is much larger than threonine and this can cause steric hindrance upon the access of 3-FBA and 4-FBA into the active site. This evidence from D469Y supports my earlier hypothesis that steric is one of the factor that controls the reaction of transketolase but the main source would arise from the side chain of D469. It should be noted that D469Y was reported to accept benzaldehyde with up to 4% isolated yield as showed in Table 4.2, this is not a significant improvement over D469T which achieved 2% isolated yield especially when taken 17-hour reaction time into account. It is possible that substituting aspartate with tyrosine could significantly reshape the active site structure which leads to a new substrate binding site. This assumption was supported by the fact that D469Y reversed the stereoselectivity of 1,3-dihydroxypentane-2-one from 3*S* to 3*R* (Smith et al., 2008; Strafford et al., 2012). However, this new binding site cannot accommodate 3-FBA and 4-FBA.

F434A was reported to give the highest yields for both benzaldehyde and 3-hydroxybenzaldehyde in the previous report with up to 10% and 6%, respectively, but it cannot catalyse 3-FBA and 4-FBA. Replacing the bulky phenylalanine side chain with alanine may enlarge the active site channel enough to accommodate benzaldehyde but this space may not be sufficient when a functional group is presence on the ring. This may explain the

lower yield in 3-hydroxybenzalde. In addition, Figure 4.2 a-d showed that the side chain of F434 is nearly perpendicular to the side chain of D469. If the binding of the carboxylic acid group to the phosphate binding residues directs the orientation of the aromatic ring as showed in Figure 4.2 b-c, the main steric hindrance may arise from the side chain of D469 rather than the side chain of F434. As a result, F434A does not facilitate the binding of 3-FBA and 4-FBA.

In order to confirm that the negative charge on the aromatic ring is essential to achieve high enzymatic activity, the bioconversion of 3-nitrobenzaldehyde, which has a weaker negative charge than 3-FBA, was investigated. Since D469T can accept 3-FBA and 4-FBA, it was used for this experiment. The chromatogram of the D469T reaction was then compared with the reaction samples that were incubated with other mutants and wild type that were found to accept neither 3-FBA nor 4-FBA (Figure 4.4). The 24-hour sample showed that the bioconversion of 3-nitrobenzaldehyde was very poor with less than 5% substrate consumption and this suggested that the partially negative charge on the nitro group may not be able to form sufficient affinity with the enzyme and lead to low activity. It is also possible that the low solubility of 3-NBA causes the slow reaction as observed with benzaldehyde and 3-hydroxybenzaldehyde.

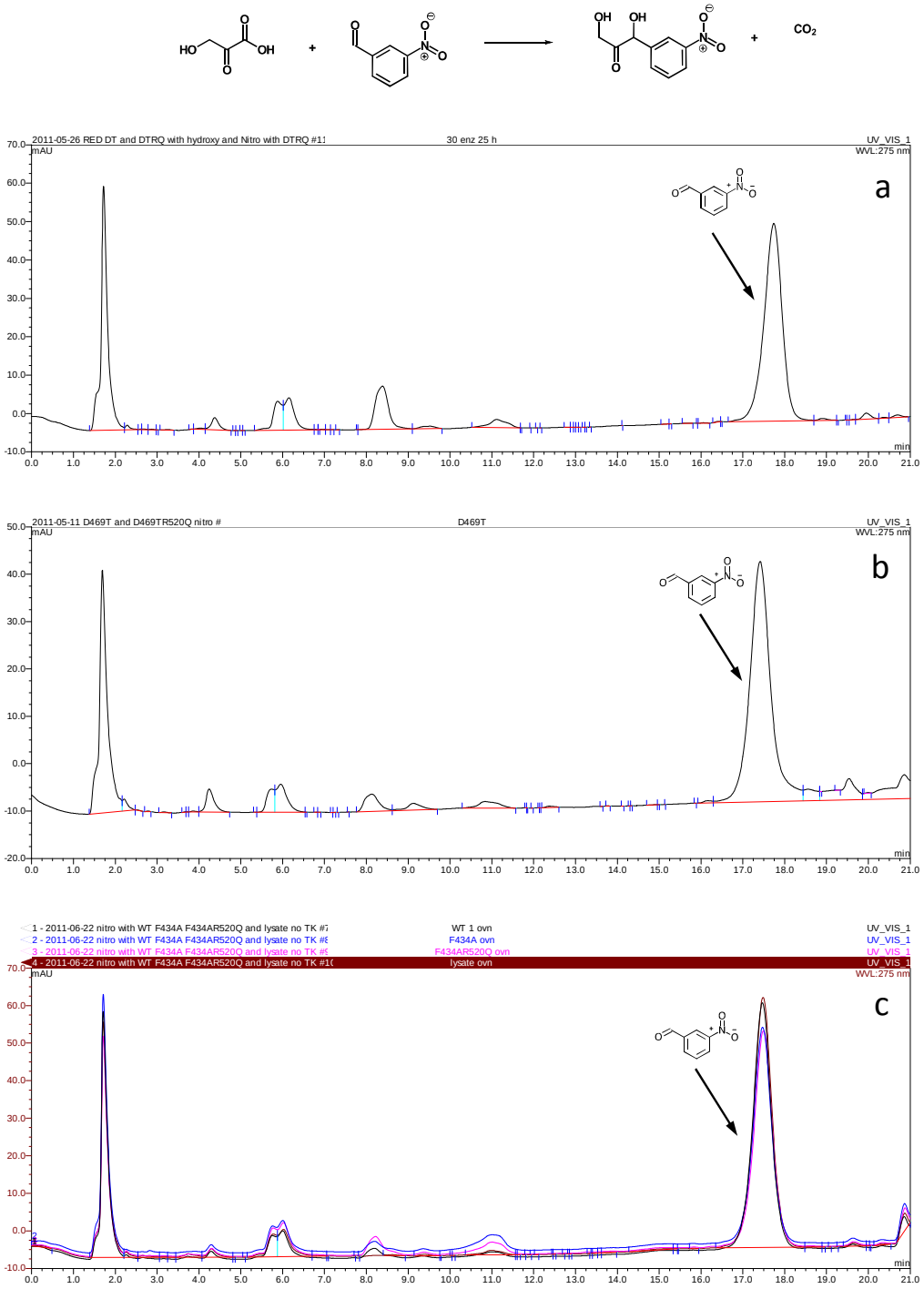


Figure 4.4 The chromatogram of 3-NBA reaction when incubated the substrate with different mutants.

a) D469T/R520Q. b) D469T. c) ●wild type, ● F434A, ● F434A/R520Q, and ● lysate with no transketolase.

Other non-negative charged aromatic substrates were tested by David Steadman (personal communication) and again no conversions were observed. These also illustrated the importance of the enzyme-substrate affinity provided by electrostatic interactions and hydrogen bonds. See the publication for further details (Payongsri et al., 2012).

4.3.2. Kinetics of the mutants towards 3-FBA and 4-FBA

The influences of the interaction between the phosphate binding residues and the carboxylic acid group of 3-FBA and 4-FBA on the improved bioconversion were investigated through site directed mutagenesis at the positively-charged phosphate binding residues which are R358, H461, and R520. Similar experiment in yeast transketolase showed that substituting these equivalent residues with alanine significantly increased the K_M values for xylulose 5-phosphate and ribose 5-phosphate (Nilsson et al., 1997). Due to these influences on the activity with phosphorylated substrates, they were the target for the enzyme-substrate affinity study. In this case, the carboxylic acid of 3-FBA and 4-FBA mimicked the phosphate group of the natural substrates.

The yields and kinetic parameters of all the mutants are showed in Table 4.3 and the Michaelis Menten plots used for deriving the kinetic constants are illustrated in Figure 4.5 and Figure 4.6. For 3-FBA, most of the mutants gave similar yields after 24 hours except a less active mutant R358L/D469T which achieved less than 10% and R358P/D469T/R520Q which gave 14% yield. The latter mutant has the highest 3-FBA affinity and could be inhibited by substrate and product to a larger extent than other mutants. Although these mutants were found to have different kinetic parameters, their yields are somewhat similar. In addition, there was up to 20% of HPA left once the product concentration reached stationary. This suggested that all these mutants suffer from a similar degree of product inhibition.

Table 4.3 The kinetic parameters of all the mutants towards 3-FBA and 4-FBA.

Mutant	Substrate	Specific activity (relative to D469T)	Conversion yield (%)	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} M^{-1})
WT	3-FBA	0	0	n.d[b]	n.d.	n.d.
D469T	3-FBA	1.0[a]	67 (1.0)	56 (10)	13 (1.5)	236 (50)
H461S/D469T	3-FBA	0.08	65 (0.6)	29 (15)	0.8 (0.2)	28 (16)
D469T/R520Q	3-FBA	0.75	67 (0.1)	13 (4)	6 (0.75)	468 (170)
H461S/D469T/R520Q	3-FBA	0.21	63 (0.7)	11 (3)	1.5 (0.2)	138 (45)
R358P/D469T	3-FBA	1.2	65 (2.5)	20 (2.5)	11.0 (0.7)	553 (80)
R358L/D469T	3-FBA	<0.1	<10	n.d.	n.d.	n.d.
R358P/D46T/R520Q	3-FBA	0.42	14 (1.3)	5.5 (1)	3.4 (0.2)	625 (115)
WT	4-FBA	0.0	0.0	n.d.	n.d.	n.d.
D469T	4-FBA	1.0[a]	30(1.7)	251 (240)	5.0 (4.4)	20 (26)
D469T/R520Q	4-FBA	0.45	13(0.4)	25 (7.6)	0.20 (0.03)	8 (2.8)
R358P/D469T	4-FBA	<0.005	1(0.1)	n.d.	n.d.	n.d.
R358L/D469T	4-FBA	0.30	5(0.1)	n.d.	n.d.	n.d.
R358P/D469T/R520Q	4-FBA	<0.001	<1	n.d.	n.d.	n.d.

Standard errors are in parentheses. The yield and specific activities of both aldehydes were determined at equimolar concentrations of HPA and aldehyde which are 50 mM for 3-FBA and 30 mM for 4-FBA. The yield and activity were determined using 10% V/V clarified lysate, 2.4 mM ThDP, 9.0 mM $MgCl_2$ in 50 mM Tris buffer, pH 7.0.

[a] The specific activity of D469T towards 3-FBA and 4-FBA are 4.6 µmol/mg/min and 0.45 µmol/mg/min respectively.

[b] n.d is not determined due to very low activity.

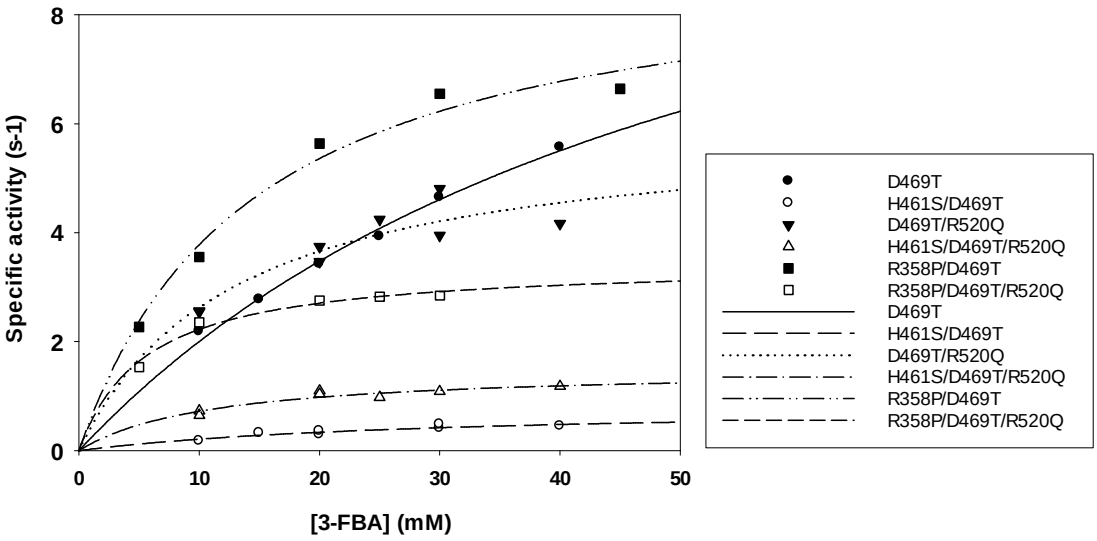


Figure 4.5 The Michaelis-Menten plot of all the six mutants for 3-FBA.

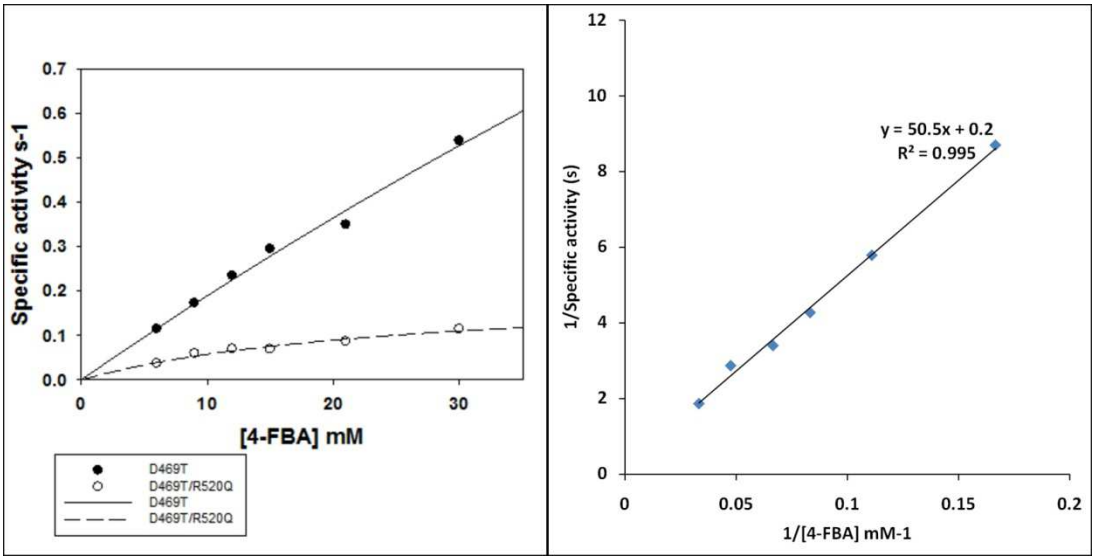


Figure 4.6 The Michaelis-Menten plot of D469T and D469T/R520Q for 4-FBA.

Lineweaver Burk plots are also shown to compare the data fitting.

The reactions of 4-FBA by all the mutants were much slower than that of 3-FBA and the conversion yields were also much lower. H461S/D469T and H461S/D469T/R520Q were far too inactive to detect any activity towards 4-FBA and were not further studied with this substrate. It is possible that in such slow reaction, enzyme inactivation, in particular by aldehyde, may be one of the key factors for low yields. In addition, we observed that HPA was degraded over the course of reaction time when the D469T containing mutants were used. When using D469T in the equimolar reaction of 3-FBA, the conversion reached 65% after 2 hours and the reaction did not progress further. The product concentration became stationary and 3-FBA remained constant but there was only 20% of HPA left and it continued to decrease. Similar trends were observed in the reaction with propionaldehyde and 4-FBA where the HPA consumption is higher than the product formation. I further investigated to see whether this degradation was catalysed by D469T or other components in the clarified lysate, buffer, background transketolase wild type, or ThDP. 50 mM HPA was incubated for 18 hours in the presence and absence of ThDP, lysate with overexpressed D469T or D469T/R520Q, wild type, XL10 Gold cells with no overexpressed TK, and 50 mM Tris as a control as showed in Table 4.4. From this table, it can be seen that the background HPA degradation occurred at 2-14 % under the presence of cell-free lysate. This degradation also occurred in Tris buffer alone. The highly expressed wild type appeared to contribute to this degradation but to a much lower degree. These results showed that other components in the reaction contribute very little effect on HPA degradation. However, only the D469T and D469T/R520Q with ThDP reactions caused greater than 90 %

HPA consumption which illustrated that the D469T mutant directly involved in HPA degradation which can influence the conversion yield especially for a long reaction time.

The kinetic study for 3-FBA showed that substituting the amino acids at the phosphate binding sites resulted in mutants with lower activity than D469T except R358P/D469T which had 20 % improvement. In this study, substrate inhibition was observed at the 3-FBA concentration above 40 mM which was excluded from further analysis. The mutation at both R358 and R520 significantly reduced the K_M values to a higher degree than the k_{cat}. Consequently, the overall the k_{cat}/K_M value of R358P/D469T and D469T/R520Q increased to 553 ± 80 s^{-1} M^{-1} and 468 ± 170 s^{-1} M^{-1}compared with 236 ± 50 s^{-1} M^{-1}in D469T. The K_M was further reduced when R358P was combined with R520Q in R358P/D469T/R520Q which resulted in a mutant with the lowest K_M value of 5.5 mM with the k_{cat} of 3.4 ±0.2 s^{-1}. With such a low K_M, this triple mutant has the highest k_{cat}/K_M with value of 625 ±115 s^{-1} M^{-1}. The k_{cat}/K_M values of the double and triple mutants are higher than that the value for glycolaldehyde by the wild type (283 s^{-1} M^{-1}) (Hibbert et al., 2007) and the D469T towards propionaldehyde (91 s^{-1} M^{-1}) (Hibbert et al., 2008). However, they are still much lower than the k_{cat}/K_M values for phosphorylated aldoses erythrose 4-phosphate ($1.7x10^6 s^{-1}$ M^{-1}) and ribose 5-phosphate ($3.1x10^5 s^{-1}$ M^{-1}) catalysed by yeast transketolase (Nilsson et al., 1998). D469T/R520Q may appear to have a better performance than D469T due to the higher k_{cat}/K_M value. This double mutant, however, has higher activities than D469T only at the concentration of 3-FBA below 30 mM. Replacing the positively charge arginine could impair the enzyme activity as observed in R358L/D469T which has less than 10% of the activity and this was far too low to study the kinetic parameters. In addition, the presence of H461S mutation in any mutant results in severely impaired enzymatic activities. The H461S/R520Q mutant has the lowest activity in all the mutants created and this was slightly recovered by the R520Q mutation in the triple mutant H461S/D469T/R520Q.

Interestingly, replacing any positively charged residue at the phosphate binding site decreased the K_M values. The alterations in the K_M values were observed in H461S/D469T and H461S/D469T/R520Q mutants where their K_M values are 29 mM and 11 mM. More significant changes in the K_M values were observed in R538P/D469T and D469T/R520Q where K_M values significantly decrease from 56 ± 10 mM in D469T to 20 ± 2.5 mM and 13 ± 4 mM and the lowest K_M value in R358P/D469T/R520Q. The fact that H461S causes a smaller change in the K_M value may imply that H461 may contribute a weaker interaction

than the two arginine residues. The interaction between H461 and the carboxylic acid group may be a weak hydrogen bond rather than an electrostatic interaction. It could be suggested that the side chain hydroxyl group of H461S may act as a hydrogen bond donor. However, the shorter side chain of serine may make this hydroxyl group out of hydrogen bond range. As a result, the lowest K_M value in R358P/D469T/R520Q is unlikely to be due to the interaction between the H461 and the carboxylic acid group on 3-FBA. It is possible that this triple mutation caused some structural rearrangements in the active site which lead to an exposure of R91 residue which could provide a strong interaction.

The decreases in the K_M values indicate an increase in enzyme-substrate affinity when the positively charge residue was removed. This somehow differs from what was observed in yeast transketolase where the mutation at the phosphate binding site leaded to a low affinity between the enzyme and phosphorylated substrates (Nilsson et al., 1997). Since higher binding affinity represents a more stabile binding mode, the difference in the trend of the K_M values between phosphorylated and carboxylated substrates implied that the two types of charge require different interactions and they were not stabilised in the same way. Several positively charged amino acids are required to stabilise the 2- phosphate group upon the binding and removing any of these residue may easily destabilise it due to the burial of charge on the phosphate group itself. In contrast, the presence of several positively charged amino acids destabilised the binding of 1- carboxylic acid possibly due to the burial of excess charge on the active site.

Table 4.4 The % HPA left after incubating in different conditions for 18 hours.

Reaction conditions	% HPA left
D469T with ThDP	5.7
D469T/R520Q with ThDP	5.7
50 mM Tris	92
50 mM Tris with ThDP	95
D469T without ThDP	98
D469T/R520Q without ThDP	89
WT with ThDP	86
WT without ThDP	92
XL10 Gold with ThDP	91
XL10 Gold without ThDP	97

Aldehyde was not added into any condition.

In addition to the effect on the K_M, all these mutations reduced the k_{cat} values which may arise from repositioning the whole 3-FBA molecule in such a way that the aldehyde group of 3-FBA was located further away from the reactive carbanion-ThDP intermediate. In R358P/D469T, D469T/R520Q, and R358P/D469T/R520Q, their k_{cat} were reduced from 13 s^{-1} to 11 s^{-1}, 6 s^{-1}, and 3.4 s^{-1}, respectively. The most significantly impaired k_{cat} mutant was found to be H461S/D469T (0.8 s^{-1}) which also has the lowest activity. This was slightly recovered in H461S/D469T/R520Q to 1.5 s^{-1}. In fact, H461S mutant and any mutants bearing H461S were found to have lower enzymatic activity towards glycolaldehyde and propionaldehyde as reported in Chapter 3. I previously suggested that this mutation causes a global instability and overall reduction in the enzymatic activity. Here, the k_{cat}values of H461S/D469T and H461S/D469T/R520Q further support my hypothesis. The instability of H461S mutant was further illustrated by the lower soluble fraction which was found to be 10 % of the total soluble cellular protein while the wild type has 16% expression level. In order to further confirm that H461S causes a general reduction in the enzymatic activity, I tested H461S/D469T and H461S/d469T/R520Q with propionaldehyde. Propionaldehyde is a very small substrate and unlikely to form a strong specific interaction with H461. The specific activities of H461S/D469T and H461S/D469T/R520Q towards propionaldehyde are 0.08 µmol/mg/min and 0.22 µmol/mg/min, respectively. These are 5.8% and 16% of the activity of D469T towards propionaldehyde. The slight improvements on the activities for both substrates upon R520Q mutation was thought to be due to the stabilising tendency of R520Q on certain D469 variants through the co-evolved network (Strafford et al., 2012).

D469T catalysed the bioconversion of 4-FBA at a specific activity of 0.45 µmol/mg/min which is approximately 10 times slower than the reaction of 3-FBA. The k_{cat} for 4-FBA is 5 s^{-1}, approximately 2.5 times lower than 3-FBA, while the K_M value towards 4-FBA is nearly five times higher than 3-FBA although both substrates carry the same charge. These leaded to the k_{cat}/K_M value of 20 $s^{-1}M^{-1}$ which is nearly 20 folds lower than the k_{cat}/K_M value for 3-FBA and this implied that 4-FBA is a much less favourable substrate than 3-FBA. Assuming that the carboxylic acid groups of both substrates bind to the active site at a similar position as suggested in Figure 4.2 b and d, the aldehydes of the two substrates are not at the same position which resulted in the significant difference in their k_{cat}. This also implied that the relative orientations of the aldehyde and the carboxylic acid group are important for a productive binding and catalysis. Although the *para* position of the carboxylic acid in 4-FBA would electronically favour nucleophilic attack at the carbonyl

group than the *meta* position in 3-FBA, the k_{cat} values between the two substrates illustrated that electronic effects may not be a major factor influencing the enzymatic activity. In addition, 3-nitro group is much more electron withdrawing than the carboxylic acid group, the enzymatic activity towards 3-nitrobenzaldehyde is much poorer than 3-FBA and 4-FBA.

The removal of any positively charged residue significantly impaired the enzymatic activity and yield towards 4-FBA. Most of the double mutants have very low activities to analyse the cause of the loss in activity except D469T/R520Q. In the case of D469T/R520Q, this was due to the 25-fold reduction in the k_{cat} rather than losing the enzyme-substrate affinity. In fact, the introduction of R520Q significantly decreased the K_M value to 25 mM which is still approximately 2 times higher than the K_M value for 3-FBA by the same mutant. R358P/D469T was found to have much lower activity towards 4-FBA than D469T and D469T/R520Q while it has higher activity towards 3-FBA than both mutants. Therefore, the improvement is only specific for 3-FBA through the improved affinity. In R358L/D469T, the activity towards 4-FBA was partially impaired and the activity towards 3-FBA was reduced to a larger extent.

The mutagenesis and kinetic studies so far illustrated that removing a positively charged amino acid resulted in a lowering ofK_M and k_{cat} for both 3-FBA and 4-FBA. However, 3-FBA was positioned in a more active orientation than 4-FBA in all double and triple mutants.

Throughout my study, the bioconversions of 3-FBA and 4-FBA at 10 mM were found to be much faster and higher than the reaction of benzaldehyde which was performed at 50 mM. This suggested that solubility of the substrate may not be an issue in controlling the rate of the reaction. It was illustrated that the replacements of negatively charged D469 to a neutral charge threonine or tyrosine have improved the enzyme-substrate affinity towards non-hydroxylated substrate propionaldehyde (Hibbert et al., 2008). The interaction between propionaldehyde and the transketolase active site would rely on the hydrophobic interaction alone. To test whether neutralising the positive charge on the active site would favour the hydrophobic interaction, D469T/R520Q was tested with benzaldehyde at 50 mM. The conversion yield was less than 3% which is not significantly improved over D469T. This suggested that hydrophobic interaction alone is insufficient to achieve high yield and activity in aromatic substrates.

4.3.3. Computational docking of 3-FBA and 4-FBA

The binding energies, distances between the carbonyl group to the enamine in D469T and D469T/R520Q were summarised in Table 4.5. The computational docking of 3-FBA into D469T showed that the carboxylic acid of 3-FBA form hydrogen bond and electrostatic interactions with R358, S385, H461, R520 as showed in Figure 4.2 b and c. 2 binding clusters of 3-FBA in D469T were predicted. In the first cluster, the carbonyl group form hydrogen bond with the side chain of H26 which placed the carbonyl group 4.9 Å away from the enamine-ThDP. In the second cluster, the carbonyl group of 3-FBA forms hydrogen bond with the main chain O-atom of H473 which places the carbonyl group 7.2Å away from the enamine-ThDP making this conformation rather unproductive due to a longer distance to the enamine.

The docking of 3-FBA into D469T/R520Q is slightly different from D469T. The whole 3-FBA molecule was moved further away from the enamine-ThDP and S385 is not forming hydrogen bond with the carboxylic group of 3-FBA but it forms the hydrogen bond with the carbonyl group instead (Figure 4.2c). This placed the aldehyde 6.6Å away from the enamine which may reflect the 2-fold lower in k_{cat} in D469T/R520Q.

4-FBA was predicted to form one binding cluster in D469T (Figure 4.2d). Unlike 3-FBA, only one of the oxygen atom on the carboxylic acid was predicted to interact with R358, H461, and R520 but not S385. This interaction network was less extensive than that of 3-FBA in D469T which may reflect the 5-fold higher K_M in 4-FBA. The aldehyde of 4-FBA was placed 5.1Å away from the enamine-ThDP which is 0.16 Å further than the aldehyde of 3-FBA. Although this further distance may suggest the 2.5-fold lower k_{cat} in 4-FBA, this distance was 1.5 Å closer than the aldehyde of 3-FBA to the enamine-ThDP in D469T/R520Q which has similar k_{cat} value to the k_{cat} of 4-FBA in D469T. Therefore, in addition to the distance between the carbonyl group to the enamine-ThDP, the relative orientation of the carbonyl group relative to the enamine intermediate may also be a crucial factor governing the reaction. Finally, 4-FBA was found to have multiple binding clusters in D469T/R520Q and these clusters were found to have similar binding energies. The distances between the carbonyl group to the enamine were 4.6 Å, 6.5 Å, 8.3 Å, and 8.9 Å with the population of 35%, 10%, 16%, and 26% respectively. The non-productive clusters could block the active site and prevent other 4-FBA molecule to bind in a productive conformation which resulted in low observed k_{cat} value.

Table 4.5 The free energy binding (ΔG), the predicted K_d and experimental K_M towards 3-FBA and 4-FBA for D469T and D469T/R520Q.

Mutant	ΔG predicted (kcal/mol)		K_d predicted (mM)		K_M experiment (mM)		Distance of aldehyde to enamine (Å)	
	3-FBA	4-FBA	3-FBA	4-FBA	3-FBA	4-FBA	3-FBA	4-FBA
D469T	-3.35	-3.15	3.30	4.64	56	251	4.9	5.1
D469T/R520Q	-1.96	-1.91[a]	35.3	38.7	13	25	6.6	4.6-8.9[b]

The K_d values were calculated from $\Delta G = -RT.\ln(K_d)$.

So far, the predicted positions of 3-FBA and 4-FBA were consistent with the experimental kinetic data. These binding energies were used to calculate the dissociation constants and compared with the wild type data as showed in Figure 4.7. When comparing the binding energies of 3-FBA and 4-FBA in D469T and D469T/R520Q (Table 4.5), only the calculated K_d from D469T were consistent with the calculated K_d for wild type. Only the data of D469T fitted well with the wild type data while the calculated K_d values from D469T/R520Q were off the trend. The deviations observed in D469T/R520Q may arise from some structural rearrangements upon energy minimisation by Charm. Although the structure of D469T was prepared in the same way, there could be more structural changes in the double mutant. This small difference may lead to the systematic changes in the energy calculation in the computational docking.

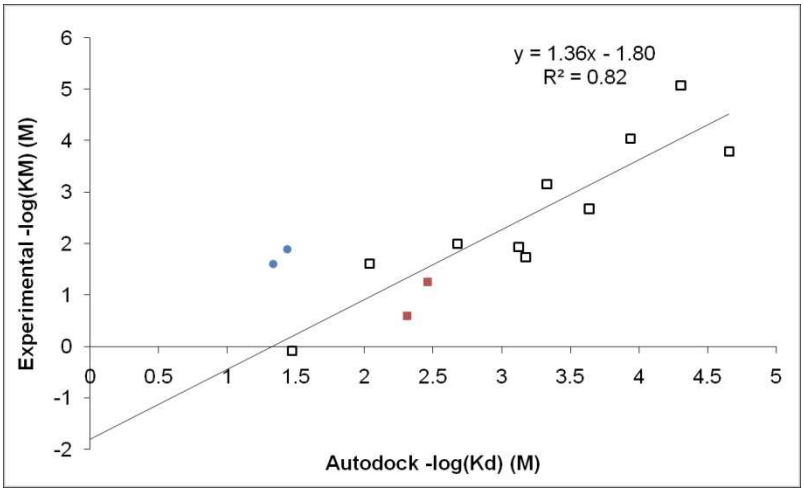


Figure 4.7 Comparison between the experimentally determined binding constants (K_M) and the calculated K_d from binding energy using AudoDOCK .

(□) *E. coli* wild-type, (■) D469T and (●) D469T/R520Q transketolases. The linear best fit was plotted using the data from the wild type only.

4.4. Conclusion

I have illustrated that the phosphate binding site at the entrance of the transketolase active site can be used to enhance the bioconversion of benzaldehyde derivatives that contain negatively charged carboxylic acid. The position of the carboxylic acid itself on the aromatic ring influences the position of the aldehyde group. This affects the catalytic activity as illustrated that the *meta* position appears to have higher activity as shown in 3-FBA than 4-FBA. The difference may be due to the relative positions of their aldehyde where the *meta* position is placed in a more productive location. In addition, replacing the positively charged residues leads to higher affinity but lower k_{cat} and hydrophobic interaction alone is insufficient for a fast conversion and high yield for aromatic aldehydes.

5. Chapter 5: Second Generation Engineering of Transketolase for Aromatic Ketodiols

5.1. Introduction

Transketolase accepts aromatic aldehydes and heteroaromatic aldehydes with poor activity and yield (Galman et al., 2010). In Chapter 4, it was found that the presence of the carboxylic group on 3-formylbenzoic acid can improve the activity up to 250 times faster than benzaldehyde with the yield up to 67% within 2 hours (Payongsri et al., 2012). These enhancements were found to be due to restoring the hydrogen bonds and electrostatic interaction between the carboxylic acid group and the phosphate binding residues of the enzyme. D469T was found to be the most active mutant towards 3-FBA and 4-FBA with the specific activities of 4.6 μmol/mg/min and 0.45 μmol/mg/min respectively. The activity of D469T for 3-FBA is actually higher than any previously reported aromatic aldehydes and non-hydroxylated aldehydes. However, the catalytic turnovers of these aromatic substrates are still much lower than natural substrates (13.2 s^{-1} and 5.0 s^{-1} for 3-FBA and 4-FBA respectively). The study in yeast transketolase revealed that, ribose 5-phosphate has a k_{cat} value of 46.3 s^{-1} (Wikner et al., 1997) and for erythrose 4-phosphate the k_{cat} value is 69 s^{-1} (Nilsson et al., 1998). These relatively low k_{cat} values suggested some possibilities for improving transketolase by protein engineering. In addition to improving the performances towards 3-FBA and 4-FBA, protein engineering is a powerful tool to further enhance the activity of transketolase towards neutral and polar aromatic aldehydes which are still poorly accepted. This is especially important for the synthesis of chloramphenicol and its derivatives where the substituents are non-charged and the synthetic potential of transketolase could be further expanded.

A number of directed evolution strategies are available with their own advantages and drawbacks. However, they tend to involve the handling of large library size which can be resource intensive. An alternative option is to construct smart libraries by targeting residues to those within active site close to the reaction centre and highly conserved residues. This strategy has been shown to improve substrate activity and also influence the stereoselectivity of the enzymes (Morley and Kazlauskas, 2005; Paramesvaran et al., 2009; Toscano et al., 2007). As a result, the size of the library is not too large to manually handle and I adopted this strategy to design my library.

The analysis of *E. coli* transketolase active site in the previous study has classified active site residues into two shells (Hibbert et al., 2007). Residues within the first shells are located within 4 Å from the substrate molecule whereas residues in the second shells are within 10 Å from the ThDP molecule (Hibbert et al., 2007). Among these residues, the active site of transketolase consists of a number of highly conserved histidine residues on the phosphate binding domain and these residues are H26, H69, H100, and H261 (Costelloe et al., 2008; Schenk et al., 1997). Mutagenesis studies in yeast transketolase on the highly conserved histidine residues showed that H30 and H263 (H26 and H261 in *E. coli*) may act as an acid/base catalyst (Wikner et al., 1997). My previous study in Chapter 3 showed that H26Y has no activity and recently, this mutant was also recreated by another research group and it was found to be inactive (Yi et al., 2012). H103 was suggested to be involved in substrate recognition and stabilisation of the intermediate (Wikner et al., 1995). H69 was identified to bind to the phosphate group of ThDP and may be involved in the proton transfer between a water molecule and C1-hydroxyl group of the donor-ThDP adduct (Wikner et al., 1997). Therefore, these histidine residues were excluded from the library constructed in this Chapter. My previous study in Chapter 4 illustrated that substituting H461 with serine may lead to enzyme instability as a lower fraction of soluble protein was observed. This mutation also globally reduced the enzyme activity, although this residue does not appear to interact with nearby amino acids (Payongsri et al., 2012). Therefore, H461 was also excluded from the library. From all these eliminations, the target residues were limited to only R358 and S385. The crystal structure of *E. coli* transketolase revealed that the side-chain hydroxyl group of S385 and the guanidinium group of R358 form hydrogen bonds with the phosphate group of xylulose 5-phosphate (PDB ID: 2R5O) and fructose 6-phosphate (PDB ID: 2R8P) (Asztalos et al., 2007). These hydrogen bonds were also observed in the crystal structure of yeast transketolase with erythrulose 4-phosphate (PDB ID: 1NGS) and replacing R358 equivalent residue in yeast transketolase significantly decreases the affinities towards xylulose 5-phosphate and ribose 5-phosphate (Nilsson et al., 1997). I have previously confirmed in Chapter 4, that R358 in *E. coli* transketolase interacts with the carboxylate group of both 3-FBA and 4-FBA from the R358P and R358L mutants (Payongsri et al., 2012). The previous saturation mutagenesis studies at both residues showed that only R358I has approximately 2 times higher activity than wild type towards both glycolaldehyde (Hibbert et al., 2007) and propionaldehyde (Hibbert et al., 2008). However, the library of S385 did not produce any mutant with improved activity towards glycolaldehyde or propionaldehyde. In this Chapter, the saturation mutagenesis of

these sites has been introduced within the D469T/R520Q mutant in order to improve the activity of transketolase towards aromatic aldehydes.

5.2. Materials and methods

5.2.1. Chemicals and reagents

All aldehyde substrates were purchased from Sigma-Aldrich (Aldrich® Chemistry, UK). HPA was synthesised according to the protocol referred in Chapter 2.

5.2.2. Transketolase library

Although D469T has the highest activities and yields for 3-FBA and 4-FBA in the last chapter, all the libraries were constructed on D469T/R520Q due to its stability which allows it to withstand further mutations and also minimise inclusion body formation (Strafford et al., 2012). The target residues were limited to those that are within the 1st shell (Hibbert et al., 2007) but excluding the catalytic histidine residues (H26, H69, H100, H261) and H461 where its mutation was shown to have an impact on enzyme stability and activity (Payongsri et al., 2012). These eliminations left only R358, S385, D469, and R520. Since D469T/R520Q was used as a template, R358 and S385 are the only two targets available.

The triple mutants of both libraries were constructed by Quickchange site directed mutagenesis according to the standard protocol. All the *dpn*I digested PCR products were transformed into XL10 Gold competent cells (Stratagene, La Jolla, California, USA). The sequences of the primers were used as listed below. The target codons are underlined and the bases to be changed are in bold letters. N represents an equal concentration of A, T, C, and G. H is 50% T and 50% A. S is 50% C and 50% G.

R358X GAAAATCGCCAGC**NHN**AAAGCGTCTCAGAATG

S385X GCTGACCTGGCGCCG**NHS**AACCTGACCCTGTGG

Q520R CACTGATCCTCTCC**GT**CAGAACCTGGCGCAG

The last primer is to reverse the triple mutants to S385Y/D469T and S385T/D469T.

5.2.3. Library Screening

According to the sequences of the primers above, both libraries were expected to produce 16 possible amino acids. The library coverage was calculated by the online tool: GLUE1AA (http://guinevere.otago.ac.nz/cgi-bin/aef/glue1AA.pl) (Patrick et al., 2003) and

picking 24 colonies was expected to give 11.6 distinct amino acids under the assumption that the libraries are diverse. 6 random colonies from each library were picked and their plasmids were sequenced in order to verify the diversity of each library.

Uneven cell lysis can influence the amount of released enzyme, hence the reaction rate. Therefore, whole cells were used in the screening of all three substrates. 24 colonies from each library were picked and inoculated into 12 ml of LB media with 150 µg/ml ampicillin. They were then cultured at 37 °C, shaking at 250 rpm for 18 hours. Their final OD600 values were measured before harvest. These cells were aliquoted into 2 ml eppendorf tubes which were then centrifuged at 17,000 g for 10 minutes to harvest and the supernatant was discarded. To each eppendorf tube containing supernatant-free cells, 200 µl of 2x cofactor solution (4.8 mM ThDP and 18 mM $MgCl_2$ in 50 mM Tris buffer, pH 7.0) was added. After the cells were fully resuspended, they were left to incubate with the cofactors for 20 minutes. 150 µL of these cell-cofactor mixtures were transferred to a borosilicate microplate (Radleys, Essex, UK). The reaction was then started by the addition of 150 µl of 2x substrate (100 mM 3-FBA and 100 mM HPA, 80 mM 4-FBA and 80 mM HPA, or 30 mM 3-HBA and 60 mM HPA). All substrates were prepared in 50 mM Tris-HCl buffer and adjusted the pH to 7.0. The plate was then seal to prevent evaporation (Thermo Scientific Nunc) and shaken at 200 rpm, 3 mm amplitude to avoid sedimentation of the cells. 3-FBA samples were taken at 1 hour and 24 hours after the reaction was started in order to screen for mutants with higher activities and yields. The higher activity mutants would assume to give higher conversion at 1 hour. My previous experiments illustrated that the reactions of 4-FBA and 3-HBA are very slow compared with 3-FBA and their conversion yields by D469T, which is the best mutant, were also low (Payongsri et al., 2012). Taking samples within 1 hour would give low signal to noise ratio. In addition, HPA degradation also competes with the ketol transfer reaction of transketolase. As a result, the mutants with higher activity tend to give higher yields. Therefore, only the samples at 18 hours were taken to quantify the final yields. This method assumes that all the mutants have similar expression levels of transketolase.

380 µl of 0.1% TFA was added to 20 µl of sample to quench the reaction. These were then centrifuged for 3 minutes at 13,300 rpm before the supernatants were analysed by HPLC according to the methods stated in Chapter 2. The conversion of 3-FBA and 4-FBA reactions were quantified by the same method as stated in Chapter 4 (Payongsri et al., 2012). Racemic dihydroxy ketone of 3-HBA, 1,3-dihydroxy-1-(3-hydroxyphenyl)propan-2-

one, was synthesised and purified according to the published protocol (Galman et al., 2010) and this was used as a standard for quantification.

5.2.4. Cell culture and protein quantification for detailed enzyme kinetics

Cell cultures of all the mutants started from streaking glycerol stock onto LB-plates containing 150 µg/ml of ampicillin and cultured according to the standard protocol. All of them were cultured at the final volume of 50 ml in LB media with 150 µg/ml of ampicillin in 250-ml shake flasks. Clarified lysate was prepared according to the protocol in Chapter 2, section 2.4.

For the 3-HBA kinetic study, cells harvested from 50 ml culture were resuspended in 2 ml of 50 mM Tris buffer, pH 7.0 before sonication. This is to produce clarified lysate with transketolase concentration at least 0.8 mg/ml. The total protein concentration was quantified by Bradford assay and the expression level of TK was quantified by SDS-PAGE and densitometry. The expression levels of transketolase mutants from both libraries are approximately 20 % of the total soluble protein in the cell-free lysate.

5.2.5. Detailed enzyme kinetics

All kinetic studies were performed in triplicate at the reaction volume of 300 µl in 1.5 ml glass vials. In all the reactions, the concentration of HPA maintained at 50 mM and the concentration of the aldehydes were varied. For 3-FBA, the concentrations were from 3-90 mM while 4-FBA the concentrations were between 6-30 mM and for 3-HBA, the concentrations were 5-25 mM. For the reaction of 3-FBA and 4-FBA, 30 µl of clarified lysate was incubated with 30 µl 10x cofactor solution and 140 µl 50 mM Tris buffer, pH 7.0 for 20 minutes. 100 µl of 3x substrate solutions were then added to the lysate-cofactor solutions to start the reaction. For 3-FBA, the reaction samples were taken every taken every 3 minutes for 15 minutes while the samples of 4-FBA reaction were taken at every 15 minutes for 90 minutes.

Due to the lower solubility of 3-HBA, the substrate stock solutions were prepared as 2x concentration. The reactions of 3-HBA were prepared by incubating 50 µl of clarified lysate with 30 µl of 10X cofactor solution and 70 µl of 50 mM Tris buffer, pH 7.0 for 20 minutes. 150 µl of 2x substrate solutions were then added to start the reaction. The samples were taken every 30 minutes for 180 minutes.

20 µl of the reaction samples were quenched by the addition of 380 µl of 0.1% TFA which were then centrifuged and analysed by HPLC as described in Chapter 2, section 2.7.2.

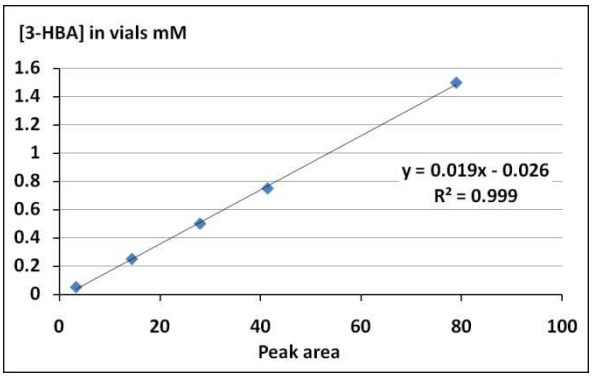


Figure 5.1 The standard graph of 3-HBA.

All standard solutions were analysed by ACE5 C18 column and detected at 275 nm as described in section 2.7.2. The peak areas were plotted against the concentration of 3-HBA in the vials.

5.2.6. Computational modelling of the binding of aldehyde substrate

The details for the preparation of the energy-minimised structures of the mutants were the same as the previously mentioned in Section 4.2.8, Chapter 4. Computational dockings were performed by David Steadman (Chemistry Department, UCL) with the same parameter as Section 4.2.8, Chapter 4 and the previous publication (Payongsri et al., 2012).

5.3. Results and discussions

5.3.1. The 3-HBA reaction samples when analysed by HPLC

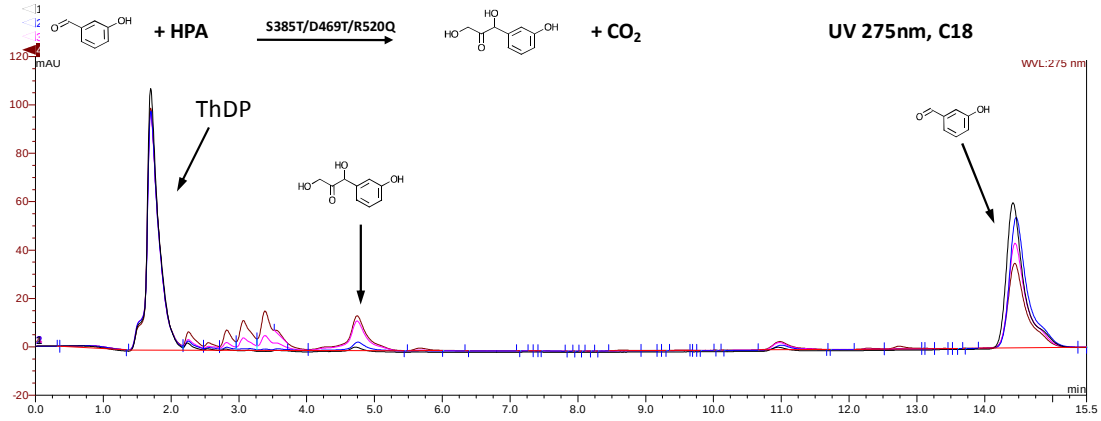


Figure 5.2 The chromatograms of the 3-HBA reaction samples.

The samples from a reaction catalysed by S385T/D469T/R520Q were taken at (●) 30, (●) 60, (●) 150 minutes, and (●) 18 hours, respectively. All the samples were analysed by ACE5 C18 column and the mobile phases used were 80% methanol and 0.2 M acetic acid as described in section 2.7.2.

5.3.2. High-throughput screening

In Chapter 4 it was found that transketolase mutants bearing the D469T mutation appeared to catalyse HPA degradation which could lead to lower conversion yields in 3-FBA and 4-FBA (see Chapter 4, Table 4.4) (Payongsri et al., 2012). This characteristic would also explain the very low yield in a much slower reaction of 3-HBA where only 4-6 % yield was observed (Galman et al., 2010). The yield also appeared to be higher as the specific activity towards that particular substrate increase. This was illustrated in the case of D469T catalysing 3-FBA and 4-FBA where 65% and 30% yields were achieved but the activity for the former substrate was 10 times higher than the latter (Payongsri et al., 2012). Therefore, it could be speculated that mutants that give high yields at 18 hours would be active mutants for 4-FBA and 3-HBA. Determining the conversion at the 18 hours would also reduce the noise. Since the reaction of 3-FBA is much faster than other two substrates, the activities of all the mutants were measured at the 1 hour where less than 35% conversion was achieved in D469T.

Prior to further analysing any mutants, the high throughput method, which used whole cells, was evaluated by comparing the conversions and yields with activities and yields determined in clarified lysate where the concentrations of transketolase were quantified. Cell-free lysate of seven newly created mutants were selected from both

libraries to quantify their specific activities towards 50 mM 3-FBA which were compared with the 1-hour conversion yield in whole cell system. Since different mutants were harvested at slightly different OD600, there would be a small variation in the transketolase concentration in the screening reaction and their 1-hour conversions would need to be standardised by their OD600. These were then plotted against the clarified lysate activities as shown in Figure 5.3. The whole cell data correlated well with those of clarified lysate which illustrated that the whole cell data can represent the activities obtained in the lysate. Only the clarified lysate activity of S385E/D469T/R520Q was found to be much higher than the whole cell system. This mutant was later found to be highly active towards 3-FBA and the deviation from the high throughput may arise from the mass transfer across the cell membrane that slowed down the whole cell catalysis.

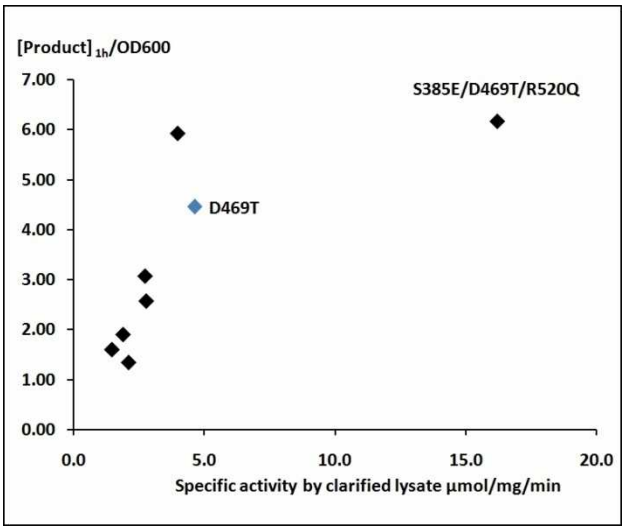


Figure 5.3 Comparison of the specific activities by clarified lysate and the 1-hour conversions in whole cell which were standardised by OD600.

For 4-FBA and 3-HBA, five mutants were selected. A good agreement between the clarified lysate and high throughput screen was illustrated by the linear relationship between the two data sets as shown in Figure 5.4. The time course sampling of both 4-FBA and 3-HBA also support my initial hypothesis that higher activity mutants gave higher 18-hour yields (see later for integrated results). However, systematically higher yields were obtained from clarified lysate which implied that mass transfer into the cells could limit the conversion yields in the whole cell system. For example, D469T gave 30% conversion yield in clarified lysate (Payongsri et al., 2012) whereas the whole cell D469T gave only 15.6 % yield after 18 hours. The conversion of 3-HBA by D469T also follows the same trend where the whole cell achieved 4.9% yield while lysate gave 6.7% (Table 5.1). The good correlation between the lysate and whole cell data allow me to further analyse all the mutants in both

libraries towards three substrates. The yields towards these substrates were also compared.

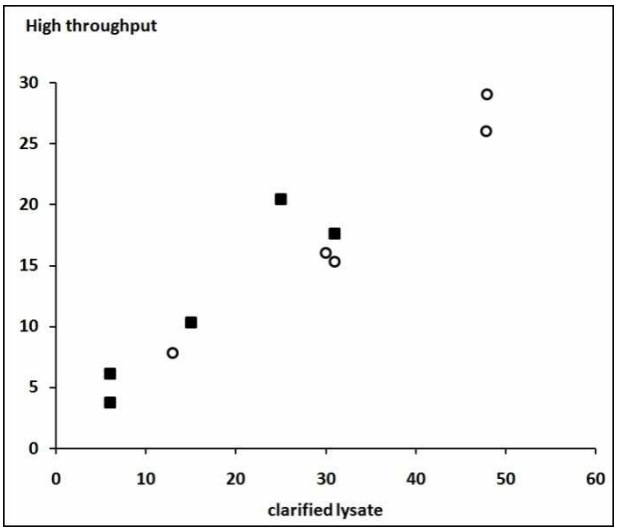


Figure 5.4 The comparison of the 18-hour yields by clarified lysate and whole cells in high throughput screen towards 4-FBA ○ and 3-HBA ■.

5.3.2.1. S385X/D469T/R520Q library screening

Although the previous studies of S385X library towards glycolaldehyde (Hibbert et al., 2007) and propionaldehyde (Hibbert et al., 2008) did not give any mutant with higher specific activity, the high throughput screening of the S385X/D469T/R520Q library showed that there were several mutants with improved activity towards the selected aromatic aldehydes (Figure 5.5). In the whole cell reaction, at least 5 mutants can achieve 2-5.5 fold higher yield than D469T for 3-HBA, 2 mutants approximately doubled the yields of 4-FBA and one mutant appeared to have nearly 2 folds higher activity than D469T for 3-FBA. The majority of the mutants still gave similar or slightly lower conversion yields for 3-FBA than D469T after 1 hour in the whole cell reaction. Interestingly, the yields of 3-HBA by the mutants in this library appeared to have a strong positive relationship with the yields towards 4-FBA. This correlation suggested that the improvement of 4-FBA and 3-HBA could employ a similar mechanism but this differed from that of 3-FBA which may involve the relief of substrate inhibition. This will be discussed in the next section.

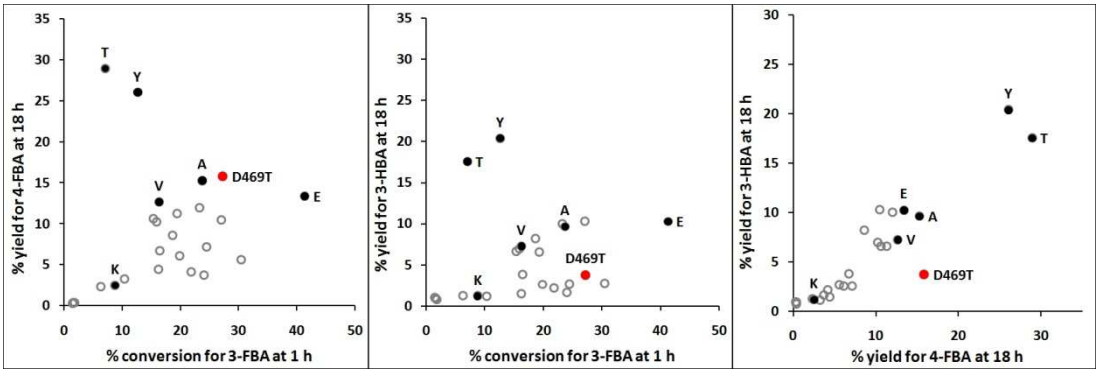


Figure 5.5 The cross comparison of the yields between three substrates by S385X/D469T/R520Q library.
The activities towards 3-FBA were measured after 1 hour where the reaction achieve less than 35% conversion. The final yields of 4-FBA and 3-HBA were measured at 18 hours. All reactions were performed in freshly harvest whole cells.

In addition, I initially speculated that HPA degradation is competing with the productive ketol transfer reaction and this dictated the final yield of aromatic aldehydes. This leads to a hypothesis that higher activity mutants will give higher final yields. The specific activities and yields for 3-HBA determined by clarified lysate under the same substrate concentrations as the high throughput screen showed that D469T had an activity of 0.0071 µmol/mg/min with the 6.7% conversion after 18 hours (Table 5.1) while S385V/D469T/R520Q,S385Y/D469T/R520Q and S385T/D469T/R520Q achieved 15%, 25%, and 31% respectively. The specific activities of these mutants by clarified lysate were 0.03, 0.071, and 0.073 µmol/mg/min, respectively which were 5-12 times higher than D469T. This positive correlation between yield and activity was also observed in the bioconversion of 4-FBA. This increase in the final yield in higher activity mutants confirmed that the ketol transfer reaction is competing with the HPA degradation. This characteristic was previously studied and identified that D469T bearing mutants, but not the background host enzyme or buffer, causes the observed HPA degradation(Payongsri et al., 2012). This relationship between yield and specific activity was, however, not obvious in the 3-FBA reaction which was partly due to substrate inhibition and relatively fast catalysis which will be discussed later.

The DNA sequencing revealed that substituting S385 with hydrophobic amino acids (V, A, Y), polar amino acid (T), and negatively charged amino acid (E) can improve the bioconversion of 3-HBA. Since rather diverse amino acids can substitute S385, it implied that these mutants had improved k_{cat} instead of providing better interactions with the hydroxyl or carboxylic acid groups of the substrates with an exception for S385E. This may be unexpected because S385 binds to the phosphate group of natural substrates.

5.3.2.2. R358X/D469T/R520Q library screening

The previous attempt at introducing saturation mutagenesis at R358 gave R358P and R358I mutants with 1.5 and 2.1 fold improvements of activity for glycolaldehyde (Hibbert et al., 2007). Both mutants also had 1.5 and 1.9 times higher activity than wild type towards propionaldehyde (Hibbert et al., 2008). In this study, however, saturation mutagenesis at this position on D469T/R520Q did not appear to yield a mutant with higher activity towards the selected aromatic aldehydes (Figure 5.6). In contrast to the S385X/D469T/R520Q library, the activity towards 3-FBA has a positive correlation with the conversion yield of 4-FBA.

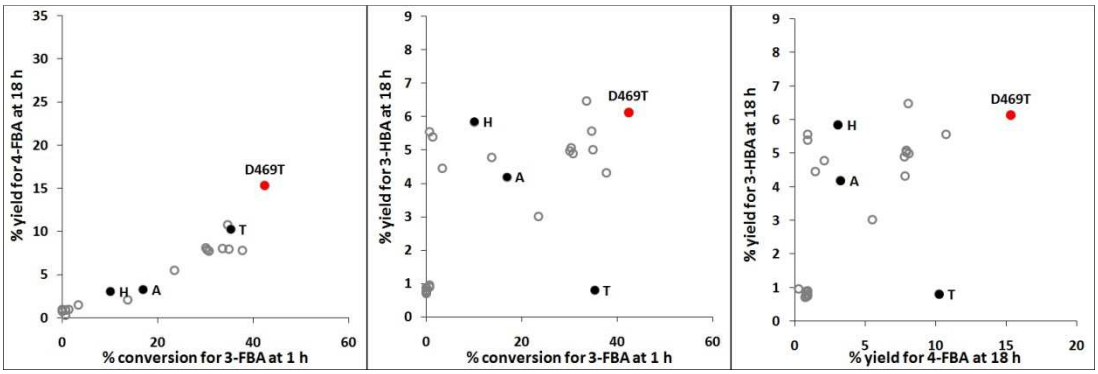


Figure 5.6 The cross comparison of the yields between three substrates by R358X/D469T/R520Q library.
The activities towards 3-FBA were measured after 1 hour where the reaction achieve less than 35% conversion. The final yields of 4-FBA and 3-HBA were measured at 18 hours. All reactions were performed in freshly harvest whole cells.

In addition to screening this library, some triple mutants obtained from random sampling were sequenced and assessed their performances using clarified lysate. R358E/D469T/R520Q and R358Q/D469T/R520Q were found to have diminished activity for 3-FBA. The activities of R358H/D469T/R520Q, and R358I/D469T/R520Q were marginally lower than D469T. The high throughput screen showed that approximately a third of the mutants from the R358X/D469T/R520Q library had 1-hour conversions of 3-FBA similar to D469T while nearly half of the library had significantly lower activities towards 3-FBA. The rest of the library had partially impaired activity towards 3-FBA. In addition, the yields of 4-FBA by the majority of the mutants were lower than 50% of D469T while more than half of the mutants from this library achieved similar yields to D469T for 3-HBA. Since I previously showed in Chapter 4 that R358 interacted with the carboxylate group of 3-FBA and 4-FBA and stabilised their bindings, it is possible that the removal of the positively charged arginine would have a detrimental effect on the bioconversion of the carboxylated

aromatic aldehydes to a greater extent than 3-HBA. This could also explain the correlation between the 3-FBA activity and 4-FBA yield obtained from this library (Figure 5.6). Besides that, it could imply that the mutation at R358 did not have a strong influence on 3-HBA compared with 3-FBA and 4-FBA. Alternatively, 3-HBA may have a rather different orientation than 3-FBA and 4-FBA. The evidence for this hypothesis arose from two mutants as follows. The R358T/D469T/R520Q mutant maintained 60% activity towards 3-FBA and 4-FBA but it had significantly lower activity towards 3-HBA. R358H/D469T/R520Q only maintained the activity towards 3-HBA but not 3-FBA or 4-FBA despite the fact that it still had a positively charged histidine. It is still surprising that none of the mutants from this library has improved activity towards any of the substrates.

The high-throughput screen of both libraries illustrated that the S385X library can tolerate a wider range of amino acid substitution than the R358X for 3-FBA and 4-FBA. Replacing R358 with other amino acids had a tendency to maintain the enzymatic activity only towards 3-HBA while the substituting S385 was more likely to improve the activity towards 3-HBA. Although both *E. coli* and yeast transketolase crystal structures illustrated that S385 forms hydrogen bonds with the phosphate group of natural substrates (Asztalos et al., 2007; Nilsson et al., 1997), it may not contribute significant interaction with the aromatic substrates. On the other hand, the R358 residue, together with H461 and R520, strongly interacted with the carboxylic acid groups of both 3-FBA and 4-FBA as shown in the previous study (Payongsri et al., 2012).

5.3.3. Kinetic parameters of the mutants with 3-FBA, 4-FBA, and 3-HBA

The high-throughput screen illustrated that none of these mutants simultaneously had higher activity towards all three substrates and all the mutants appeared to have lower activity towards at least one substrate. This apparent shift in substrate specificity can be investigated in detail by kinetic studies which can reveal whether this is due to a loss in enzyme-substrate affinity or a reduction in the catalytic activity k_{cat}. Therefore, the mutants with higher activities towards at least one substrate were further analysed for their kinetic parameters in order to gain an insight into their mechanisms on enhancing the activity towards each substrate as well as the causes of the shift in substrate specificity. The specific activities of all three mutants towards different aldehyde concentrations were plotted in order to find the k_{cat} and K_M of each mutants for different aldehydes as shown in Figure 5.7. The details for the calculation are the same as those in Chapter 4, section4.2.5, Equation 4.1. These kinetic parameters are summarised in Table 5.1.

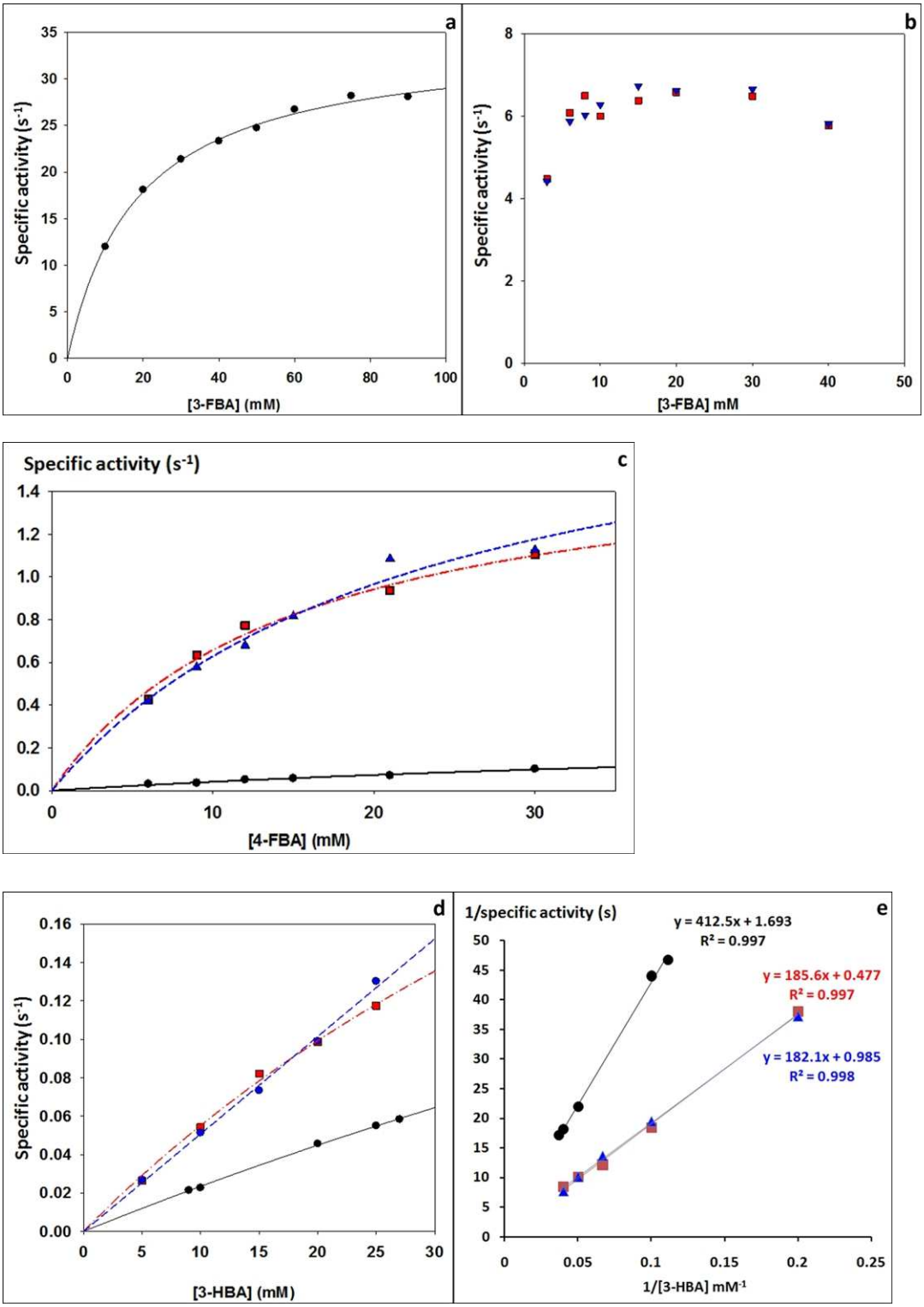


Figure 5.7 The kinetic plots of all three mutants for 3-FBA, 4-FBA, and 3-HBA.

The specific activities of all the mutants were plotted against different aldehyde concentrations. In all figures, ● is S385E/D469T/R520Q, ■ is S385Y/D469T/R520Q, and ▲ is S385T/D469T/R520Q. For 3-HBA, Lineweaver Burk plot is also provided due to the linearity of the data.

Among all of my mutants from both libraries, only S385E/D469T/R520Q was found to have higher activity towards 3-FBA than D469T. The specific activity of this mutant at 50

mM 3-FBA was 3.5 folds higher than that of D469T. Besides the increase in the k_{cat} value to 34.3 ± 1.6 s^{-1}, 3-FBA inhibition was not observed up to 90 mM of 3-FBA although the K_M was 18.3 ± 2.6 mM which was 3 times lower than D469T. This level of affinity was similar to when R520Q was introduced in D469T/R520Q. As a result, the major factor that improves the performance of this mutant was the k_{cat}. The significant improvement in the k_{cat} led to a 4-fold improvement of k_{cat}/K_M value over D469T to 1900 ± 270 s^{-1} M^{-1}.

My previously constructed mutants were inhibited by 3-FBA at 40 mM (Payongsri et al., 2012) while this substrate inhibition was relieved in S385E/D469T/R520Q. Transketolase follows a ping-pong mechanism (Schenk et al., 1998) where HPA has to bind to the active site to form dihydroxyethyl-ThDP (DHE-ThDP) intermediate before the binding of aldehyde. The inhibitory effect of aldehyde may arise from a more preferential binding of aldehyde over HPA and the catalysis of HPA is not rapid enough. In order to identify whether these are the cases, the kinetic parameter of S385E/D469T/R520Q towards HPA was also investigated and listed in Table 5.1. The K_M value of S385E/D469T/R520Q towards HPA was found to be 52 ± 4 mM which was nearly 3 times higher than the K_M value towards 3-FBA. Since this K_M value suggested that the binding of HPA is less preferential the 3-FBA, kinetic parameter alone could not explain the relieved inhibitory effect of S385E/D469T/R520Q. It is possible that there are some structural rearrangements within the active site of S385E/D469T/R520Q in such a way that a new 3-FBA binding site is formed and does not compete with the binding of HPA. This hypothesis was investigated by structural modelling and computational docking in the later section.

In addition to 3-FBA, S385E/D469T/R520Q also had 6-fold higher activity towards 3-HBA than D469T with 15% yield which was 2.2-fold higher. The replacement of a carboxylic group for the hydroxyl group in 3-HBA led to a large K_M value of 245 mM. The k_{cat} towards 3-HBA was found to be 0.59 s^{-1} which was much lower than that of 3-FBA. Although the overall performance of this mutant over D469T, the activity for 3-HBA was rather moderate. Due to the fact that D469T and D469T/R520Q had very poor activities for 3-HBA, their kinetic parameters could not be obtained for comparison. The activity of this mutant for 4-FBA was only 19% of D469T which was even lower than D469T/R520Q. The k_{cat} of this mutant towards 4-FBA was found to be 0.3 ± 0.1 s^{-1} which was similar to D469T/R520Q while the K_M value was 72 mM, nearly 3-fold higher than D469T/R520Q. Consequently, the overall performance is very low.

In Table 5.1 below, the highest k_{cat}/K_M value belongs to the reaction of 3-FBA catalysed by S385Y/D469T/R520Q. The k_{cat} value of this mutant towards 3-FBA was 7.0 ± 0.3 s^{-1} which was slightly less than half of D469T and the K_M was only 1.3 mM, 43 times smaller than that of D469T and nearly 10 folds lower than D469T/R520Q. This mutant experienced substrate inhibition at 20 mM which was much lower than what has been observed in D469T and D469T/R520Q and this could possibly be due to its very strong affinity with the substrate. S385T/D469T/R520Q also possessed similar kinetic parameters where its $k_{cat}$and $K_M$valueswere 7.3 ± 0.2 s^{-1} and 1.7 mM. However, both mutant achieved greater than 50% yield for 3-FBA. My previously reported R358P/D469T/R520Q mutant also had very low K_M value but substrate inhibition occurred at the concentration above 40 mM while only 14% yield was achieved (Payongsri et al., 2012). This comparison illustrated that the low K_M value alone may not be able to explain the cause or the relief mechanism of substrate inhibition. This also supported my hypothesis above that S385E/D469T/R520Q may acquire other mechanisms to elevated substrate inhibition and this can be through providing a new 3-FBA binding site through structural rearrangement.

Due to the very high k_{cat}/K_M value of S385Y/D469T/R520Q, I would like to evaluate the applicability of this mutant in organic synthesis and also large scale synthesis by comparing its performance with other mutants. If this mutant and D469T were not inhibited by 3-FBA, the activity of this mutant at a concentration above 65 mM would still be lower than D469T. S385E/D469T/R520Q, on the other hand, had a lower k_{cat}/K_M than S385Y/D469T/R520Q but due to the very high k_{cat}value, this mutant has the higher activity at any 3-FBA concentration. This comparison illustrates that although S385Y/D469T/R520Q has a higher k_{cat}/K_Mvalue, it is not an appropriate mutant to be used. For industrial purposes where space-time yield has to be considered, S385E/D469T/R520Q would be the most attractive.

S385T/D469T/R520Q and S385Y/D469T/R520Q were the only two mutants that their specific activities towards 4-FBA were greater than doubled of D469T. My previous study revealed that substitution R520 or R358 for non-charged amino acids led to a significant reduction in the activity towards 4-FBA and the k_{cat} of D469T/R520Q was 25 times lower than that of D469T (Payongsri et al., 2012). This kinetic study illustrated that S385T/D469T/R520Q and S385Y/D469T/R520Q recovered the k_{cat} from 0.2 s^{-1} in D469T/R520Q to 2.1 s^{-1} and 1.7 s^{-1}, respectively while their K_M values are not significantly altered from D469T/R520Q. The combination of the restored k_{cat} and the lower K_M resulted

in an overall higher enzymatic activity towards 4-FBA. Besides that, the k_{cat}/K_M values of both mutants were 4.5-5 times higher than that of D469T, and over 10 folds when compared with D469T/R520Q.

All the kinetic studies so far illustrated that the mutation at S385 can alter the k_{cat} for both 3-FBA and 4-FBA and one mutant may have very different k_{cat} values towards both substrate. Given than that the carboxylate group of both 3-FBA and 4-FBA interacts with R358, the relative position of the aldehyde of both 3-FBA and 4-FBA in all three mutants would be different. These suggested that the enzymatic activity towards aromatic aldehyde does depend on the type and the position of the substituent group on the aromatic ring relative to the aldehyde which in turn affects the position of the aldehyde relative to the catalytic centre. The alteration enzymatic activity upon mutation towards certain aromatic substrates is not due to the change in the general catalytic activity but rather a shift in substrate orientation and proximity to the reactive centre.

My previous study on 3-FBA and 4-FBA indirectly illustrated the importance of the enzyme-substrate affinity but none of the mutants could be studied with non-charged aromatic aldehyde due to a very low activity of D469T and D469T/R520Q (Payongsri et al., 2012). Here is the first time that the kinetic parameters could be determined. In addition to 4-FBA, S385T/D469T/R520Q and S385Y/D469T/R520Q also had considerably higher activities and yields towards 3-HBA than D469T. The k_{cat} values of S385Y/D469T/R520Q and S385T/d469T/R520Q were 2.09 s^{-1} and 1.01 s^{-1} which were similar to their k_{cat} values towards 4-FBA but up to 4 folds lower than the k_{cat} for 3-FBA. Their K_M towards 3-HBA, however, were 100-300 times higher than the K_M for 3-FBA, and 8-25 times higher than K_M for 4-FBA. These very large K_M values led to the k_{cat}/K_M values of approximately 5.5 s^{-1} M^{-1} which was 1000 times lower than the k_{cat}/K_M for 3-FBA. These kinetic parameters from both mutants supported my previous hypothesis that the affinity between transketolase and aromatic aldehyde is one of the major factors governing the biocatalytic conversion of aromatic aldehydes (Payongsri et al., 2012). These K_M values were also much higher than the concentration of 3-HBA within the reaction and such high K_M values gave rise to a very low specific activity towards 3-HBA compared with 3-FBA and 4-FBA.

Table 5.1 The kinetic parameters of all the mutants towards 3-FBA, 4-FBA, and 3-HBA.

	Substrate	Specific activity (relative to D469T)	Yield (%)	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} M^{-1})
D469T	3-FBA	1.00[a]	67 (1.0)[b]	56 (10)	13.2 (1.5)	236 (50)
D469T/R520Q	3-FBA	0.75	67 (0.1)[b]	13 (4)	6.0 (0.75)	468 (170)
S385E/D469T/R520Q	3-FBA	3.5	63 (1.0)	18 (3)	34.3 (1.6)	1900 (270)
S385Y/D469T/R520Q	3-FBA	0.47	53 (0.8)	1.3 (0.4)	7.0 (0.3)	5400 (1490)
S385T/D469T/R520Q	3-FBA	0.52	59 (0.5)	1.7 (0.25)	7.3 (0.2)	4255 (620)
S385Y/D469T	3-FBA		41 (1.1)	n.d.	n.d.	n.d.
S385T/D469T	3-FBA		63 (1.4)	23 (10)	9 (2)	394 (185)
S385E/D469T/R520Q	HPA			52 (4)	46 (2)	880 (70)
D469T	4-FBA	1.00[a]	30(1.7) [b]	251 (240)	5.0 (4.4)	20 (26)
D469T/R520Q	4-FBA	0.45	13(0.4) [b]	25 (8)	0.20 (0.03)	8.1 (2.8)
S385E/D469T/R520Q	4-FBA	0.19	<5	72 (30)	0.3 (0.1)	4.6 (2.4)
S385Y/D469T/R520Q	4-FBA	3.38	48 (0.4)	15 (3)	1.7(0.1)	109 (21)
S385T/D469T/R520Q	4-FBA	2.09	48 (1.7)	23 (6)	2.1(0.3)	90 (26)
S385Y/D469T	4-FBA	1.49	35 (0.3)	12.1 (0.5)	1.13 (0.02)	94 (4)
S385T/D469T	4-FBA	<0.2	<5	>>30		
D469T	3-HBA	1.00[a]	6.7[c]	n.d.[d]	n.d.	n.d.
D469T/R520Q	3-HBA	n.d.	n.d.	n.d.	n.d.	n.d.
S385E/D469T/R520Q	3-HBA	6.0	15	245	0.59	2.42
S385Y/D469T/R520Q	3-HBA	12.4	25	387	2.09	5.40
S385T/D469T/R520Q	3-HBA	12.7	31	184	1.01	5.50
S385Y/D469T	3-HBA	5.5		122	0.56	4.58
S385T/D469T	3-HBA	<0.8		n.d.	n.d.	n.d.

[a] The specific activities of D469T determined at 50 mM 3-FBA/50 mM HPA, 30 mM 4-FBA/50 mM HPA, and 15 mM 3-HBA/30 mM HPA are 4.63 µmol/mg/min, 0.45 µmol/mg/ min, and 0.0071 µmol/mg/ min, respectively. [b] the kinetic data of D469T and D469T/R520Q were from Chapter 4,Table 4.3(Payongsri et al., 2012).[c] Yields were determined at 18 hours.[d] values were not determined (n.d.) due to very low activity.

So far, I illustrated that mutations can improve the enzymatic activity through 2 routes which can be either through improving enzyme-substrate affinity or the k_{cat}. If the concentration of substrate is much lower than the K_M value, lowering K_M will lead to higher catalytic activity. S385Y/D469T/R520Q and S385T/D469T/R520Q are the example of this case where their K_M values towards 4-FBA are 10 to 16 times lower than that of D469T. However, once the substrate concentration is much higher than the K_M, improving the k_{cat} will enhance the enzymatic activity. This is illustrated by the catalysis of 3-FBA where S385E/D469T/R520Q has similar K_M to D469T/R520Q but much higher k_{cat}.

5.3.4. The relationship between the enzyme activity and the final yields of different aromatic aldehydes

I previously mentioned that the low bioconversion rate was associated with low conversion yield in the mutant bearing D469T, possibly due to the competition with HPA degradation. In order to justify at which point HPA degradation has an influence on the conversion yield, the specific activities of the mutants towards each substrate were plotted against the percentage yields as shown in Figure 5.8. This plot clearly illustrated that as the specific activity increases, the percentage yield is improved which is similar to what I speculated earlier. However, the two parameters become independent when the specific activity is greater than 2.5 μmol/mg/min which is normally observed in the catalysis of 3-FBA by most mutants. Due to the fact that several mutants were strongly inhibited by 3-FBA, some mutants did not fit the trend. H461S containing mutants also globally lost the activity and did not follow the pattern. It should be noted that S385E/D469T/R520Q has much higher activity than D469T for 3-FBA but the yield did not improve. The 2-hour sample showed that there was HPA left in the reaction but the ketodiol product concentration becomes stationary which suggested that it is probably inhibited by the product to a similar degree as D469T. The catalysis of HPA degradation could also be faster in S385E/D469T/R520Q. The time course sampling revealed that HPA consumption is still faster than that of 3-FBA which supports the latter characteristics.

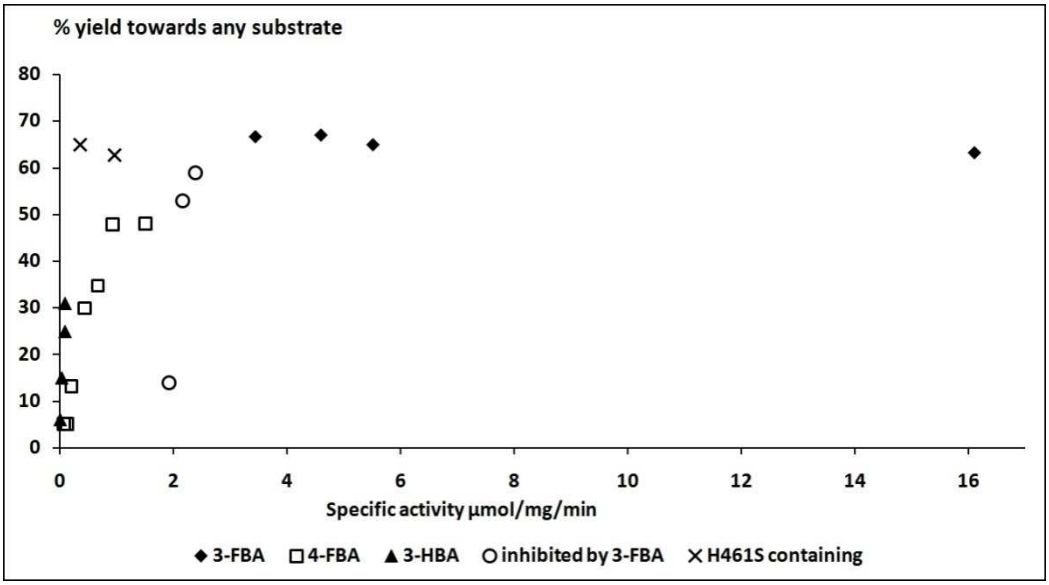


Figure 5.8 The relationship between the activities of all the mutants and the % yields of all the aromatic substrates.

5.3.5. Computational docking of substrates into the energy minimised S385E/D469T/R520Q and S385Y/D469T/R520Q structures

The screening of S385X/D469T/R520Q library revealed that several mutants maintained their activities towards a least one substrate. The kinetic study also illustrated that the mutation at S385 had a rather diverse effect on the k_{cat} of the enzyme. These evidences suggested that each amino acid substitution has different catalytic effects for all three substrates rather than a general change in the catalytic activity. These may arise from the alteration in the substrate binding orientations or the position of the aldehyde relative to the reactive enamine when the mutation was introduced. In order to rationalise this, computational docking of substrates into the energy minimised structures of the mutants was employed.

The kinetic study of S385E/D469T/R520Q alone could not explain the relief of substrate inhibition and the increase in k_{cat} towards 3-FBA. I speculated that phenomenon may arise from structural rearrangement together with a new 3-FBA binding site. The energy minimised S385E/D469T/R520Q structure was prepared according to the standard protocol (Payongsri et al., 2012). The structures of DHE-ThDP and 3-FBA were then docked into the active site. The computational docking predicted two binding clusters with different binding energy and population. The first cluster showed in green structure in Figure 5.9a had a lower binding energy together with a lower population (38%). It anchored to phosphate binding residues similar to the previous prediction of 3-FBA in D469T (Figure

5.9b) and D469T/R520Q (Figure 5.9c). In the previous study, 3-FBA was predicted to form two binding clusters in both D469T and D469T/R520Q where one cluster is a 180° rotation of the other (Figure 5.9b and c). In D469T, one cluster placed the aldehyde 4.9Å away from the enamine while the other was 7.2 Å which made it rather less active. Unlike D469T and D469T/R520Q, the binding of 3-FBA to H461 and R358 in S385E/D469T/R520Q appeared to have only unproductive conformations as the aldehyde groups were 7.2to 9.0 Å away from enamine. The second cluster (yellow stick in Figure 5.9a) was more populated (62%) and catalytically productive. It employed a new binding site as carboxylic acid cluster formed hydrogen bonds with the side chain of R91, S24, and H26, and the backbone NH group from G25. Only one conformation was predicted because the aldehyde group was anchored to the side chain of H26 and H261 via hydrogen bonds. This placed the aldehyde 4.6Å away from the enamine which is slightly shorter than that of D469T (4.9Å) (Payongsri et al., 2012). Since the aldehyde was held in placed by the hydrogen bonds, the aldehyde would be more likely to react with the enamine compared with the freely rotating aldehyde in D469T. This, together with shorter distance, could contribute to the improved k_{cat}. Furthermore, since the productive and non-productive conformations did not employ the same binding site and the productive conformation is also more populated, this could be the mechanism that substrate inhibition was relieved in S385E/D469T/R520Q.

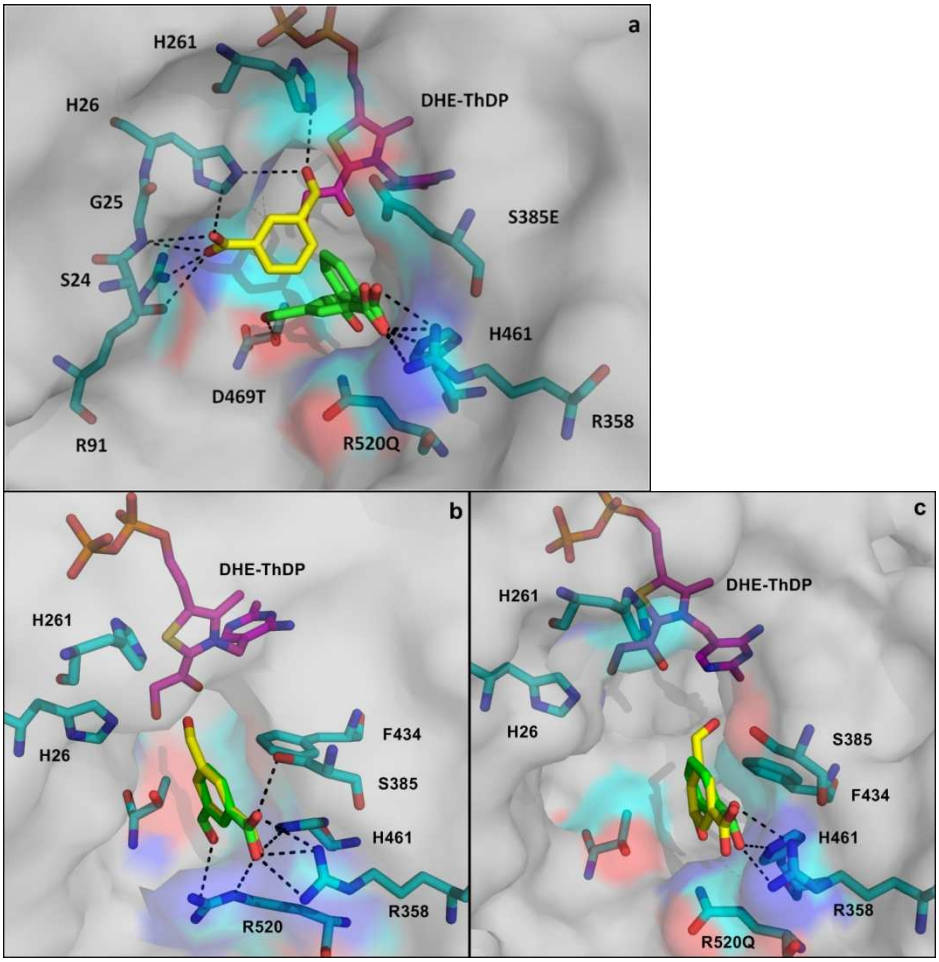


Figure 5.9 The computational docking of 3-FBA into the energy minimised a) S385E/D469T/R520Q, b) D469T, c) D469T/R520Q.

In all figure, the productive conformations are in yellow while the unproductive ones are in green. a), the 3-FBA in the first binding cluster is in green while the second cluster is in yellow. DHE-ThDP is in magenta. Hydrogen bonds are in black dotted lines.

S385Y/D469T/R520Q and S385T/D469T/R520Q had higher activity for both 4-FBA and 3-HBA. The k_{cat} values of both mutants for 4-FBA and 3-HBA are actually very similar but the reactions of 3-HBA are much slower due to the very large K_M. Computational docking of both substrates into the structures of the mutants could further give an insight into the similarities and differences between the two substrates. Here, only S385Y/D469T/R520Q was chosen to study the computational docking of 4-FBA and 3-HBA because the side chain of tyrosine is much more different from serine than threonine. The location and projection of the tyrosine side chain may be different from the side chain of serine in the wild type so that this large aromatic side chain does not block the access of the aromatic molecule. The accessibility of the aromatic aldehyde in the energy minimised

S385Y/D469T/R520Q upon computational docking would also assess the validity and reliability of the computational docking.

Unlike the side chain of S385 in wild type, the side chain of tyrosine in the energy minimised S385Y/D469T/R520Q structure did not project into the active site cavity. The computational docking of 4-FBA into this triple mutant predicted only one cluster whereas multiple clusters were found in D469T/R520Q which has very poor k_{cat} towards 4-FBA. This binding cluster in the triple mutant still employed the interaction between the carboxylic acid group with R358 and H461 (Figure 5.10a). In addition, the benzene ring of 4-FBA can form pi stacking with the aromatic ring of S385Y. These interactions hold 4-FBA in a location that placed the aldehyde group 5.1Å away from the enamine and this distance is also similar to the prediction in D469T 5.1Å (Payongsri et al., 2012). This similar distance together with only one predicted cluster could be the explanation for the similar k_{cat} to D469T.

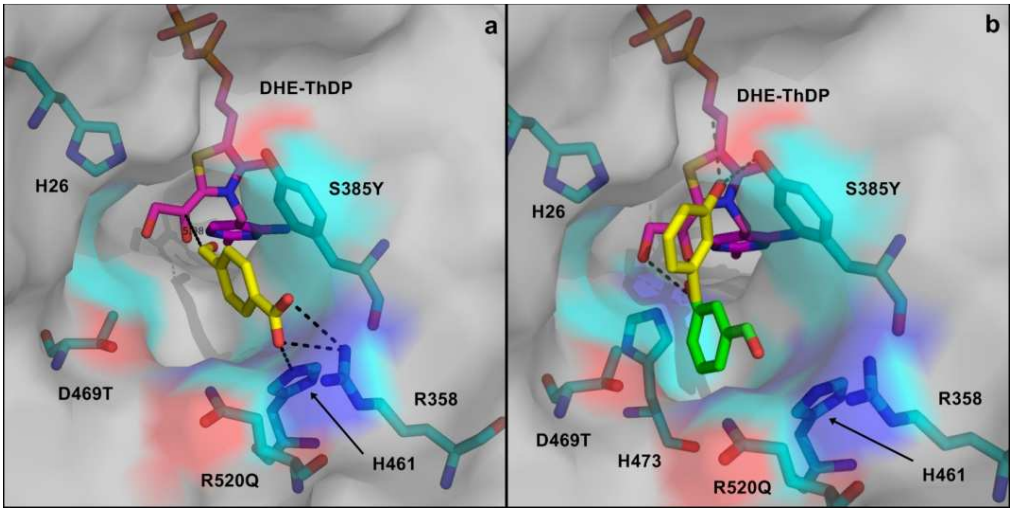


Figure 5.10 The computational docking of a) 4-FBA, and b) 3-HBA into the energy minimised S385Y/D469T/R520Q.

The catalytically active conformations in yellow. Hydrogen bonds are in black dotted lines.

The computational docking of 3-HBA in S385Y/D469T/R520Q predicted four major clusters but only two clusters showed in yellow and green in Figure 5.10b were predicted to be in the active site. However, they did not form pi stacking as 4-FBA did. The first cluster, showed as yellow stick in Figure 5.10b, formed a hydrogen bond with the side chain of S385Y through its hydroxyl group. This placed the aldehyde 4.6Å away from the enamine which is similar distance to that of 4-FBA in this triple mutant and may be the explanation for the similar k_{cat} between 4-FBA and 3-HBA by this mutant. The second cluster, as shown

in green stick, was catalytically inactive because the aldehyde group is pointing away from the active site. None of the clusters was found to interact with R358 which supported my early finding that the mutation at R358 tend to have very low influence on the activity for 3-HBA.

The reason for multiple clusters of 3-HBA could be the very low affinity between the enzyme and 3-HBA itself and hydrophobic interaction through pi stacking is probably insufficient for computational docking either. It should be noted that the computational docking of 4-FBA into the active site of D469T gave only one cluster (Payongsri et al., 2012)despite the fact that the K_M was 251 mM which is the same order of magnitude as the K_M between 3-HBA and S385Y/D469T/R520Q.

5.3.6. Interactions between S385 and R520 mutations

The kinetic studies of the two triple mutants S385Y/D469T/R520Q and S385T/D469T/R520Q revealed that they carry very similar characteristics towards all three substrates. In my previous study with 4-FBA, the replacing one of the phosphate binding arginine (R358 or R520) with a neutral amino acid led to a severely impaired k_{cat}. This was recovered by introducing either S385Y or S385T into D469T/R520Q. However, the side chains of threonine and tyrosine are very different in size and polarity. This raised a question whether both triple mutants follow the same path along of adaptive walk to reach the same optimum. Therefore, the double mutant S385Y/D469T and S385T/D469T were constructed and their kinetic parameters towards all three substrates were assessed. In order to identify the type of synergic interaction between S385T, S385Y and R520Q, the changes in the ΔG of binding and transition state from that of D469T were calculated and these were represented as ΔΔG (Mildvan et al., 1992). The ΔΔG calculated from the k_{cat} is associated with the change in the energy barrier or the stability of the transition state whereas the ΔΔG calculated from the K_M is associated with the stability of substrate binding. Therefore, a mutant with a negative ΔΔG value has a more stable substrate binding or transition state than D469T. The coupling energy ($\Delta G_{(I)}$) was calculated as below.

$$\Delta\Delta G_{(X,Y)} = \Delta\Delta G_{(X)} + \Delta\Delta G_{(Y)} + \Delta G_{(I)}$$

$$\Delta G_{(I)} = \Delta\Delta G_{(X,Y)} - \Delta\Delta G_{(X)} - \Delta\Delta G_{(Y)}$$

$\Delta\Delta G_{(X)}$ represents the changes in the free energy from the wild type upon the mutation at x position.

$\Delta\Delta G_{(X,Y)}$ is the change in the free energy from the wild type in the double mutant XY.

$\Delta\Delta G_X = \Delta G_{D469T} - \Delta G_X$

$= RT \ln(K_X/K_{D469T})$

$\Delta G_{(I)} = RT \ln \{(K_{X,Y} K_{D469T})/(K_X K_Y)\}$

K is either the K_M or k_{cat}. In all cases, D469T is the pseudo wild type.

S385T/D469T was found to be active only towards 3-FBA. The k_{cat} value was 9.0 s^{-1} which is slightly lower than that of D469T but higher than that of D469T/R520Q. Therefore, both S385T and R520Q destabilised the transition state or increased the energy barrier from D469T which resulted in the decreases in the k_{cat} value. Consequently, both double mutants had positive $\Delta\Delta G_{transition\ state}$ values. The summation of the destabilising effect from both double mutants is less than the destabilising effect in the triple mutant which made the $\Delta G_{(I)}$ negative. This means S385T and R520Q displayed partially additive effect on increasing the energy barrier for the transition state. The K_M of S385T/D469T towards 3-FBA was 23 mM which was between that of D469T and D469T/R520Q. This reduction in the K_M value was observed upon mutation of all phosphate binding residues (Payongsri et al., 2012). However, this K_M value was similar to that of H461S/D469T which suggested that S385 may not strongly participate in the interaction with the carboxylic acid group of 3-FBA. The $\Delta\Delta G_{binding}$ of the triple mutant is greater than the summation of the $\Delta\Delta G_{binding}$ of the double mutants which illustrates that S385T and R520Q further stabilise substrate binding than the simple additive combination of the effects of S385T and R520Q alone. Therefore, the stabilising effect on the binding from S385T and R520Q exhibited positive synergy.

S385T/D469T had a very low activity towards both 4-FBA and 3-HBA. As a result, the kinetic parameters towards both substrates cannot be determined. This also illustrated that S385T and R520Q must be introduced into D469T simultaneously in order to improve the activity towards 4-FBA and 3-HBA.

S385Y/D469T, on the other hand, maintained good activities towards all three substrates. However, 3-FBA inhibition was observed at even lower concentration than its triple mutant and the K_M was estimated to be lower than 10 mM. The kinetic study at lower 3-FBA concentration cannot be performed due to the detection limit of the HPLC. These evidences suggested that the K_M of the triple mutant was between the K_M of S385Y/D469T

and D469T/R520Q. The triple mutant did not stabilise 3-FBA binding further than the combination of the two mutants which represents a partially additive synergy between S385Y and R520Q. Due to the fact that 3-FBA inhibition in S385Y/D469T started at 10 mM, the k_{cat} cannot be determined. However, the specific activity of S385Y/D469T at 15 mM 3-FBA was somewhat similar to that of D469T which suggests that the k_{cat} was not impaired.

The specific activity of S385Y/D469T towards 4-FBA was 1.49 fold higher than that of D469T. This was largely due to the significant reduction in the K_M to 12.1 mM from 251 mM in D469T although the k_{cat} value of this mutant for 4-FBA was nearly 4.4 times lower than that of D469T. Both S385Y and R520Q dramatically increased the affinity for 4-FBA together with reducing the k_{cat} in the double mutants. The k_{cat} of the triple mutant was improved from the double mutants and this characteristic may suggest an epistatic or antagonistic interaction between S385Y and R520Q. However, the recovery of the k_{cat} in the triple mutant was still less than that of D469T which makes the coupling energy for the k_{cat} negative. Consequently, the destabilisation of the transition state from both S385Y and R520Q was a partially additive effect. Both S385Y and R520Q improved the binding stability from D469T. However, the absolute summation of the stabilising effect of S385Y and R520Q was greater than the stabilising effect in the triple mutant which made the coupling energy positive. Therefore, the stabilising effects from both mutations were partially additive. S385Y/D469T did not only have higher activity towards 4-FBA but also 3-HBA. The k_{cat} of the double mutant was 4-fold lower than that of the triple mutant while the K_M of the double mutant towards 3-HBA was 122 mM. Among the mutants that their kinetics can be studied, this was the lowest K_M value towards 3-HBA. Since the kinetic parameter of D469T cannot be determined, it is impossible to identify the interaction between S385Y and R520Q in the reaction of 3-HBA.

Here, I have illustrated that S385Y can improve the activity towards 4-FBA and 3-HBA independently to R520Q. Therefore, S385Y/D469T/R520Q and S385T/D469T/R520Q follow different paths to achieve a similar characteristic.

Table 5.2 The kinetic parameters and the changes in the ΔG from D469T (ΔΔG) upon the introduction of other mutations.

Mutant	Substrate	K_M (mM)	$\Delta\Delta G_{binding}$ (kJ mol^{-1})	k_{cat} (s^{-1})	$\Delta\Delta G_{transition\ state}$ (kJ mol^{-1})
D469T	3-FBA	56	0.00	13.2	0
D469T/R520Q		13	-3.60	6.0	1.93
S385T/D469T		23	-2.18	9	0.89
S385T/D469T/R520Q		1.7	-8.57	7.3	1.46
$\Delta G_{(I)}$(kJ mol^{-1})			-2.79		-1.36
D469T	4-FBA	251	0.00	5.0	0
D469T/R520Q		25	-5.68	0.20	7.87
S385Y/D469T		12.1	-7.44	1.13	3.62
S385Y/D469T/R520Q		15	-6.88	1.7	2.68
$\Delta G_{(I)}$(kJ mol^{-1})			6.24		-8.81

5.4. Conclusion

Site directed mutagenesis at the residues close to the reactive centre has improved the specific activity of transketolase towards negatively-charged and polar aromatic aldehydes. The kinetic study revealed that mutation at S385 can improve the activity via restoring the k_{cat} or improving the affinity. Computational docking was able to provide some insights into how the mutation at S385 influenced the kinetic behaviours of the enzyme for these substrates.A triple mutant, S385E/D469T/R520Q was found to have the highest k_{cat} for a non-natural substrate that has ever been reported in transketolase. The three triple mutants were found to have 5 to 12-fold improvements for the activity of 3-HBA which were high enough to determine the kinetic parameters for 3-HBA for the first time. The kinetic study of hydrophobic aromatic aldehyde 3-HBA could be done for the first time and it also supported my previous hypothesis that the affinity between the enzyme and the aromatic substrates was one of the most important factors that controls the reaction rate. The study of the synergy between S385Y/T and R520Q illustrated that different mechanisms were employed by S385Y/D469T/R520Q and S385Y/D469T/R520Q to reach similar enzymatic performance. This synergy also implied that random mutation at R520 on S385T/D469T and S385Y/D469T may provide additional beneficial mutants.

6. Chapter 6: Synthesis of novel aromatic aminodiols through the coupling of transketolase with transaminase

6.1. Introduction

Chiral amines and amino acids are the building blocks of several natural and synthetic compounds, some of which are important in pharmaceutical, agrochemical, (Höhne and Bornscheuer, 2009; Mutti et al., 2011; Rudat et al., 2012; Shin and Kim, 2001; Tufvesson et al., 2011). Several production routes of amino acids and chiral amines are available (Breuer et al., 2004). However, chemosynthesis of chiral amines are still challenging and involve toxic chemicals (Savile et al., 2010; Tufvesson et al., 2011). An alternative method is to use enzymes which becomes increasingly attractive with a range of enzymes available including hydrolases, oxidoreductases, and transferases (Höhne and Bornscheuer, 2009).

Transaminases (TAms) are transferase enzymes that belong to pyridoxal phosphate dependent enzyme family. These enzymes catalyse oxidative deamination of a donor and reductive amination of an acceptor. The reaction mechanism of transaminase is illustrated in Figure 6.1 below.

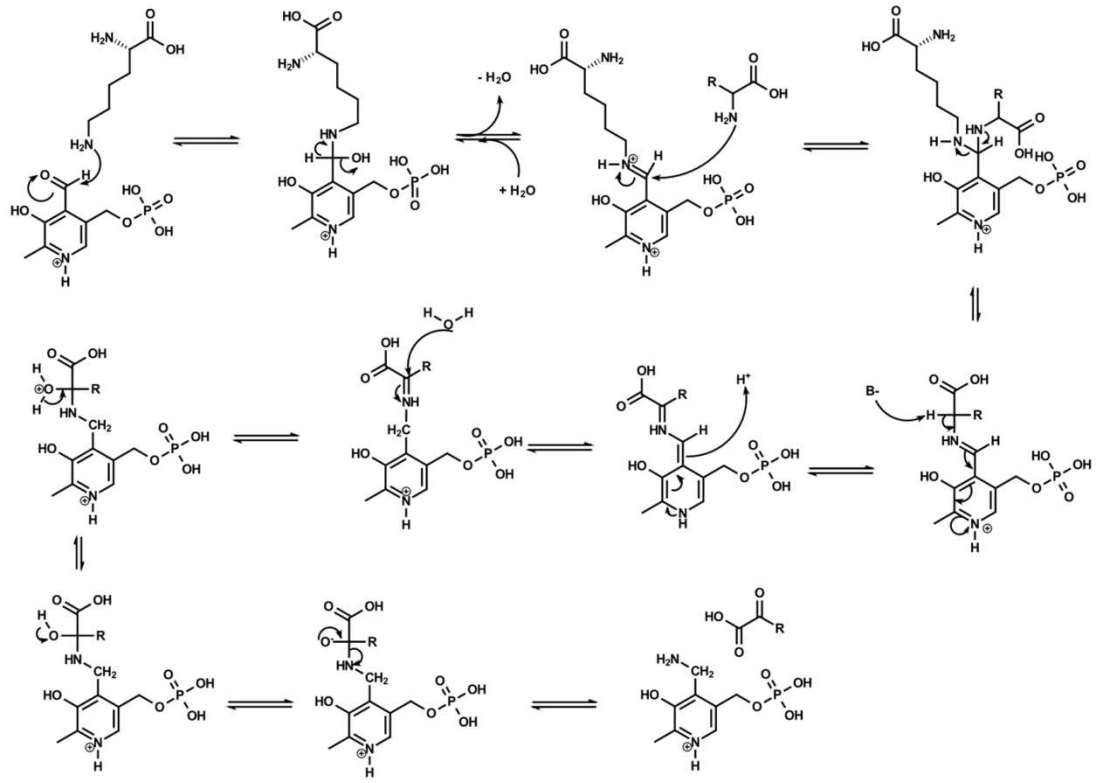


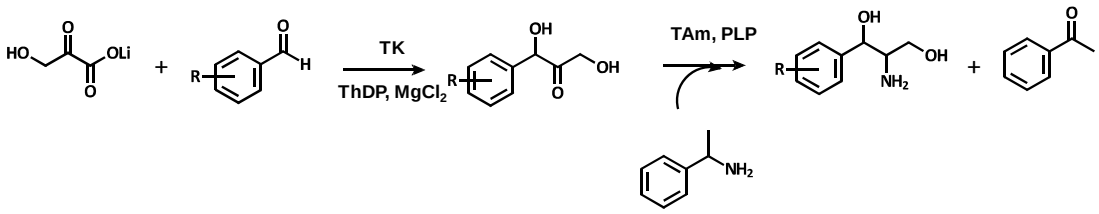
Figure 6.1 The reaction mechanism of transaminase enzymes when catalysing the transfer of an amine group from one amino acid to an alpha keto acid.

They have gained interests for the synthesis of chiral amines either by kinetic resolution and asymmetric synthesis (Koszelewski et al., 2010). The multiple sequence alignment by Protein family database (Pfam) classified transaminases into 6 classes (class I to IV) (Rudat et al., 2012; Ward and Wohlgemuth, 2010). In addition, transaminases were broadly classified according to the substrate and position of carbonyl group that is aminated (Ward and Wohlgemuth, 2010; Yonaha et al., 1983, 1976). Alpha, beta, and gamma transaminases transfer an amine group to a carbonyl group that is located at the alpha, beta, and gamma position relative to the carboxylic acid group in the substrates, respectively. ω transaminases belong to class III of the PLP-dependent enzyme sub-family and can transfer the amine group from a primary amine to aldehydes and ketones without the requirement of the carboxylic acid group (Rudat et al., 2012; Ward and Wohlgemuth, 2010). As a result, ω-transaminases have a much wider substrate spectrum than other classes of transaminases which makes them attractive in organic synthesis. Besides that, ω-transaminases also have rapid turnover, high stability (Park et al., 2012; Shin et al., 2003), their catalytic routes are more environmentally friendly (Ward and Wohlgemuth, 2010), and they do not have a requirement for cofactor recycling (Rudat et al., 2012). However, one major obstacle of using ω-transaminases is that their equilibrium constant for the

amino transferring reaction is close to 1 in most cases which means that their conversion yields are only 50% (Shin and Kim, 2001).

Chiral amino alcohols are a functionality that is found in a number of natural compounds including antibiotics such as chloramphenicols and derivatives, sphingolipids, and glycosidase inhibitors (Hailes et al., 2010; Smith et al., 2010). The chemical synthesis of chloramphenicol (Controulis et al., 1949) and thiamphenicol (Nagabhushan et al., 2000) are multi-step reactions whereas the synthesis of florphenicol requires thaimphenicol as starting material (Nagabhushan et al., 2000). Developing a shorter and simpler alternative method could provide numerous advantages over the traditional one.

Here, my research aim is to enzymatically produce novel aromatic aminodiols which can be used as building blocks in the synthesis of chloramphenicol analogues. A sequential reaction starts with the reaction of transketolase (TK) to produce aromatic 1,3-diol compound. In the next step, a transaminase catalyses the transfer of an amine group from an amine donor to the carbonyl group of the 1,3-diol as illustrated in Scheme 6.1 below.



Scheme 6.1 The reaction scheme for the production of chloramphenicol amine analogues using TK and TAm

This novel enzymatic route for the synthesis of aminoalcohols was reported in the synthesis of 2-amino-1,3,4-butanetriol (ABT) which is a functionality found in Nelfinavir (Ingram et al., 2007). This concept was further studied and developed in a number of aspects. Their kinetic constants towards the production of ABT were available for reaction modelling which aids the process design (Chen et al., 2006). A number of transketolase mutants were constructed and screened with improvements for wider range of substrates (Galman et al., 2010; Hibbert et al., 2008, 2007; Payongsri et al., 2012). Consequently, numerous amino alcohols can be produced. Our research group at UCL has also optimised aliphatic amino alcohol synthesis through transketolase-transaminase coupling system (Rios-Solis et al., 2011; Smith et al., 2010). In addition to chloramphenicol and thiamphenicol, yeast TK can accept halogenated pyruvate derivative as a donor (Esakova et al., 2009). If *E. coli* TK does the same as well as accepting an aromatic aldehyde, Florfenicol

and other flouroanalogues can be synthesised. Alongside the transketolase research, the recently discovered ω-transaminase from *Chromobacterium violaceum* DSM30191 (CV2025) was also found to accept large aromatic substrates (Kaulmann et al., 2007). Therefore, the transketolase-transaminase coupling system could have a large synthetic potential.

For the synthesis of chloramphenicol analogues, I have previously illustrated that engineered transketolase mutants accept 3-FBA and 4-FBA and produce novel aromatic dihydroxy ketones. In recent years, numerous ω-transaminases were discovered and found to have different substrate preferences (Koszelewski et al., 2010). This work assessed a small range of ω-transaminases for their ability to aminate aromatic dihydroxy ketones, although the main focus was on CV2025. This was because the substrate spectrum of CV2025 has been thoroughly investigated and aromatic substrates are more preferable. In addition, it was the only enzyme that was reported to accept 1,3-dihydroxy-1-phenylpropane-2-one (DPP) with high stereoselectivity (Smithies et al., 2009). Although this is S selective transaminase, protein engineering to reverse the stereoselectivity would be feasible especially due to the availability of the crystal structure of CV2025 (PDB ID: 4A6T)(Humble et al., 2012). In addition, the key residues that can reverse and improve the selectivity on certain aromatic substrates have been identified through rational design and 3D structure analysis (Humble et al., 2012). The enantioselectivity can be increased by a single mutation W60C and can be reversed by double mutation F88A/A231F (Humble et al., 2012).

6.2. Materials and methods

6.2.1. Materials

All the substrates for transaminase were freshly prepared in 50 mM Tris buffer and adjusted the pH to 7.5. A 40 mM PLP stock solution was prepared in 50 mM Tris buffer, pH 7.5 and stored in 80-µl aliquots at -20 °C. The concentration of the TAm in the lysate was adjusted by diluting with 50 mM Tris buffer with 0.4 mM PLP, 7.5. This solution was kept at 4 °C. Kanamycin was filtered sterile and stored in the same way as ampicillin but the final concentration was 30 mg/ml. The working concentration of kanamycin was 30 µg/ml.

6.2.2. Optimising the synthesis of the aromatic dihydroxy ketones by TK

The reactions of 3-FBA and 4-FBA were driven to completion by addition of HPA to the reaction. Both reactions started with 45 mM of HPA and 45 mM aldehyde which was prepared by addition of 2X substrate solution (90 mM HPA, 90 mM FBA in 50 mM Tris buffer, pH 7.0) into the TK lysate that has been incubated with cofactor for 20 minutes. 750 mM HPA in 50 mM Tris buffer, pH 7.0 was added to the reaction to reach the final concentration of 22.5 mM. HPA was added to the reaction when the formation of the dihydroxy ketone and the consumption of the aldehyde slowed down. All substrates were prepared in glass vials. All the reactions were performed in 1.5-ml vials and closed with snap caps.

S385E/D469T/R520Q displays the highest activity towards 3-FBA. Our previous experiment on the stereoselectivity of different mutants for 3-FBA showed that S385E/D469T/R520Q gave a racemic mixture of the 3-(1,3-Dihydroxy-2-oxopropyl)benzoic acid (3-DOPBA) (Steadman, et al, unpublished data). It was demonstrated that CV2025 can accept racemic DPP (Kaulmann et al., 2007; Smithies et al., 2009). However, the presence of the substituent group on the ring may influence the stereospecificity. As a result, S385E/D469T/R520Q was used to produce a racemic 3-DOPBA. HPA was added at 1 hour after the reaction was started and the reaction was then left overnight at 22 °C.

For 4-FBA reaction, S385Y/D469T/R520Q and S385T/D469T/R520Q were used due to their high yield and activity. 3 hours after the reaction was started, HPA was added. This was maintained at 22 °C for another 6 hours and then reduced to 4 °C for 12 hours.

6.2.3. Transaminase enzyme preparation

All the strains were provided by Prof John Ward (through Maria F Villegas-Torres). The CV2025 mutants were provided by Dawid Deszcz. The strain of TAm used in the screening are CV2025, pQR1006, pQR958, pQR1019, pQR1021, and CV2025 mutant Y153M/S/F.

A glycerol stock of each strain was streaked on LB plates with kanamycin at a concentration of 30 μg/ml. Single colonies were picked and inoculated into 250-ml flasks containing 50 ml of LB broth with kanamycin. This was cultured overnight at 37 °C, shaking at 250 rpm. These cultures were then transferred into 2-L flasks containing 450 ml of LB media with kanamycin. They were then cultured at 37 °C, shaking at 250 rpm for 3 hours. The temperature was then dropped to 30 °C prior induction by addition of IPTG to the final

concentration of 1 mM. They were cultured at 30 °C, shaking at 250 rpm for another 5 hours. 50 mg of pyridoxal-5'-phosphate was added to each culture 30 minutes before harvesting. The culture was aliquoted into 4 of 50-ml falcon tubes and centrifuged at 4000 rpm for 10 minutes. The supernatant was discarded and another 50 ml-of culture was added into these 4 tubes which were then centrifuged for another 10 minutes. This was repeated until all the culture was harvested. Each falcon tube contains cells from approximately 125 ml of culture.

Cell pellets were resuspended in 4 ml of 50 mM Tris buffer, pH 7.5 with 0.4 mM PLP. The resuspended cells were sonicated on iced according to the standard protocol. These were then centrifuged at 13,000 rpm for 15 minutes at 4 °C. The supernatant containing TAms can be used fresh or stored at -80 °C in 120-µl aliquots to be used within 1 month.

For the purification of TAm, see section 2.15.2.

6.2.4. Protein quantification

The total protein concentration within the clarified lysate was quantified by Bradford assay. The percentage of TAm within the lysate was analysed by SDS-PAGE gel and densitometry according to the protocol using BSA as a protein standard due to unavailability of purified with known concentration of transaminase.

6.2.5. HPLC

All the compounds in this study were quantified by reverse phase column ACE5 C18 according to the protocol. All samples were centrifuged at 13,000 rpm for 3 minutes before loading into the HPLC. The retention times of *S*-methylbenzylamine (MBA), 3-aminomethylbenzoic acid, 4-aminomethylbenzoic acid, and acetophenone (AP) are 3.59, 2.55, 2.26, and 7.8 minutes, respectively (see Figure 6.5 and Figure 6.6 for example). It should be noted that MBA has a much lower absorption at nearly all wavelengths than other aromatic compounds used in this study. Its absorption at 275 nm is also very low and it is recommended to use either 210 or 250 nm for the detection and quantification of MBA. An example chromatogram is shown in Figure 6.2.

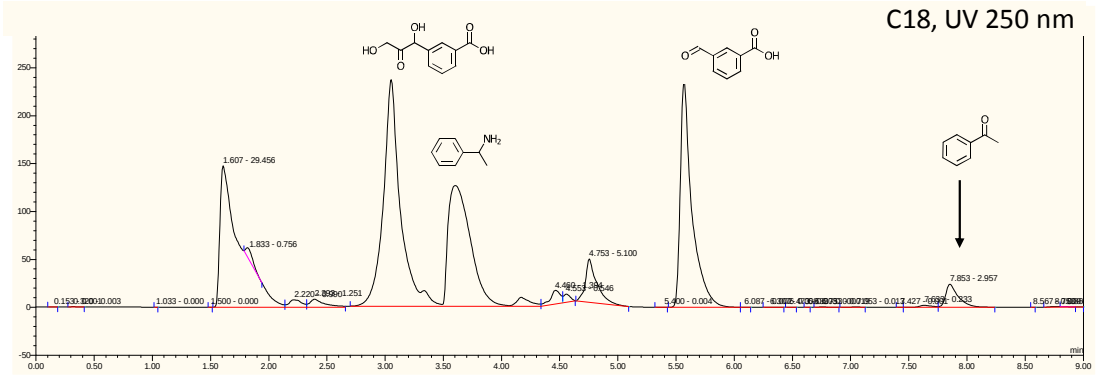


Figure 6.2 Chromatogram of the 1-minute reaction sample from the amination reaction of 3-DOPBA obtained from transketolase reaction.

The reaction consisted of 10 mM 3-DOPBA, 40 mM MBA, and less than 1 mM 3-FBA. The sample was analysed by ACE5 C18 column and detected by UV absorption at 250 nm.

6.2.6. MBA inhibition study

Throughout this study, *S*-MBA was used as the amino donor to thermodynamically drive the reaction forward due to the formation of acetophenone (Tufvesson et al., 2011). Higher concentration of MBA could also increase the reaction rate as well as shifts the equilibrium. However, it can also cause substrate inhibition. Therefore, the maximum concentration has to be determined.

The initial rates of the 3-FBA-MBA reactions where compared when using 20 mM and 40 mM MBA. Both reactions were performed in triplicate. 68 µl of clarified lysate CV2025, together with 6 µl of 40 mM PLP and 76 µl of 50 mM Tris buffer, pH 7.5, was added to 1.5-ml glass vials. The reaction was started by addition of 150 µl of the 2X substrate stock solution (40 mM 3FBA with 40 mM MBA or 80 mM MBA in 50 mM Tris buffer, pH 7.5). The vials were quickly closed by snap caps. 20 µl of samples were taken at 3, 10, 80, 120, 150, and 180 minutes and quenched by added into 380 µl of 0.1% TFA. These samples were centrifuged and the supernatants were analysed by HPLC. The peak areas of 3-aminomethylbenzoic acid were used directly to compare the initial rates.

6.2.7. Screening transaminases for aromatic dihydroxy ketones

The concentrations of the aromatic dihydroxy ketones were kept constant at 10 mM while the concentration of MBA was maintained at 40 mM.

Different ω-TAms were screened against DPP and the 3-DOPBA to see which TAm consume the highest amount of dihydroxy ketone and MBA as well as the generation of AP.

The TAms used are CV2025, pQR1006, pQR958, pQR1019, pQR1021, and CV2025 mutant Y153M/S/F.

The 450-µl reaction was prepared by adding 200 µl of TAms lysate, 9 µl of 40 mM PLP, and 90 µl of 200 mM MBA to a 1.5-ml vial. The reaction was then started by the addition of 151 µl of optimised TK reaction sample into the TAm mixture. The vial was closed by snap cap and incubated at 22 °C by Thermomixer Comfort (eppendorf, UK). 75 µl of samples were added into 225 µl of 0.1% TFA to quench the reaction. This was then centrifuged at 13,000 rpm for 3 minutes and the supernatant was analysed by the HPLC. The samples were taken at 1 minute, 2 hours, 6 hours, and 26 hours. The screening for the 4-DOPBA was conducted in the same way but only CV2025 was used.

6.2.7.1. Temperature and pH optimisation

All substrate stock solutions were prepared in 50 mM Tris buffer and adjusted the final pH to 7.5 and 8.5. Clarified lysate were prepared by resuspending cell pellets in 50 mM Tris buffer, pH 7.5 and 8.5 with 0.4 mM PLP. All different pH reactions were performed at 22 °C, 30 °C, 37 °C, and 45 °C which were maintained by Thermomixer Comfort (Eppendorf, UK). A control reaction was set up by mixing crude TK reaction with TAm lysate, PLP to the same final concentration as the TAm reaction but MBA was not added. This was done in order to compare the degree of dihydroxy ketone rearrangement over the period of time.

6.2.7.2. Control reactions

Several control reactions were set up in order to confirm that the growing peak is associated with the utilisation of the aromatic dihydroxy ketones and to assess the background acetophenone production. The reactions using cell-free and his-tag purifiedCV2025 were compared in order to compare whether the presence of cellular metabolites causes the side reactions or the increase in the area of certain peaks on the chromatogram. Clarified lysate was prepared in 50 mM Tris and HEPES, pH 7.5 in order to find out whether the background reactions were due to the buffer used. Both purified and clarified lysate CV2025 were incubated with and without MBA in both buffers. To confirm that Tris or HEPES does not catalyse the reaction, two sets of reactions were set up where the enzyme was substituted with Tris and HEPES and they were incubated under the same conditions. 20 µl of samples were taken at 1 minute, 3 hours, and 20 hours to be analysed by HPLC. All experiments were assessed without the addition of known amino acceptors and all the reactions were performed at 30 °C.

Clarified CV2025 was prepared in 50 mM Tris buffer pH 7.5 and 50 mM HEPES pH 7.5 as above procedure. Less than 1-week old, his-tag purified CV2025 was provided by Amanatuzzakiah Abdul Halim. 200 mM MBA was prepared in 50 mM Tris and HEPES and adjusted the pH to 7.5. 40 mM PLP was prepared in 50 mM Tris buffer, pH 7.5.

6.2.8. Competitive reaction

In order to find out whether the 4-DOPBA could access into the active site of CV2025, a competitive reaction between the 4-DOPBA and 4-FBA itself was performed. Due to the fact that 4-DOPBA was synthesised using clarified lysate, other components within the lysate may influence the activity of CV2025. Therefore, the 4-FBA amination rate was also compared with the presence and absence of TK lysate and ThDP.

The 4-DOPBA was enzymatically synthesised from 45 mM reaction with 10% clarified lysate at 500-µl scale. 50 µl of S385T/D469T/R520Q was incubated with 50 µl of 10X cofactor solution and 150 µl of 50 mM Tris buffer, pH 7.0 for 20 minutes before the addition of 250 µl 2X substrate solution (90 mM HPA, 90 mM 4-FBA in 50 mM Tris buffer, pH 7.0). The reaction temperature was maintained at 22 °C for 18 hours which gives 50 % conversion making the leftover 4-FBA concentration of 22.5 mM. All the concentrations at the end of the reaction were confirmed by HPLC. The left over 4-FBA concentration was then used as a starting substrate concentration in the TAm reaction. In parallel, 50 µl TK lysate was mixed with 50 µl of 10X cofactor solution and 400 µl of 50 mM Tris buffer, pH 7.0. This TK mixture was incubated at 18 hours at 22 °C as the reaction and was used in the second reaction below.

The final volumes of all three TAm reactions were 450 µl containing 10 µl of CV2025 clarified lysate, and 9 µl of 40 mM PLP. All the reactions had 10 mM of 4-FBA and 40 mM MBA at the beginning. In the first reaction, 200 µl of the 4-FBA/4-DOPBA reaction was added to the TAm-PLP mixture. In the second reaction, 56.6 µl of 90 mM 4-FBA and 200 µl of TK-ThDP mixture were added to the reaction. In the last reaction, only 56.6 µl of 90 mM 4-FBA was added to the TAm. 50 mM Tris buffer with 0.4 mM PLP at pH 7.5 was added to all these solutions to reach the final volume of 360 µl. All three reactions were started by adding 90 µl of 200 mM MBA to the reaction. All the reactions were done in duplicate. The 20-µl samples were taken at every 1 minute and quenched in 380 µl 0.1% TFA before analysed by HPLC. Standard graph of 4-aminomethylbenzoic acid was used for the quantification in the reaction.

6.2.9. Kinetic study of 3-FBA and 4-FBA with CV2025

The concentration of MBA in the reaction was kept constant at 50 mM. The concentrations of 3-FBA and 4-FBA in the reaction were varied from 3-15 mM. The enzyme concentration in the reaction was 0.18 mg/ml. The reactions were done in triplicate at the final volume of 300 µl.

50 µl of clarified CV2025 lysate, 6 µl of 40 mM PLP, and 94 µl of 50 mM Tris buffer pH 7.5 were added to the 1.5-ml vials. The reaction was started by adding 150 µl of 2X substrate solution to the reactions at every 15-second interval and the vials were quickly closed with snap cap. 20 µl of the reaction were taken and quenched in 380 µl 0.1% TFA before analysed by the HPLC according to the standard protocol. The samples of 3-FBA reaction were taken every 10 minutes while the samples of 4-FBA reactions were taken every 1 minute. Commercially available of 4-aminomethylbenzoic acid was used to prepare standard concentration for the quantification of this compound in the reaction.

For 3-FBA reaction, the rate of 3-FBA consumption was used to calculate the initial rate. Standard 3-FBA was prepared for quantification.

6.3. Results and discussions

6.3.1. The yield of aromatic dihydroxy ketones after HPA fed-batch

All the aromatic substrates that I used so far with transketolase mutants never reached 100 % conversion when the reaction was started with equimolar concentrations of HPA and aldehyde. If the dihydroxy ketone from the reaction was used directly without purification steps, the leftover aldehyde could compete with the dihydroxy ketone in the TAm reaction step. Since CV2025 was reported to catalyse aromatic aldehydes at a much higher rate than DPP (Kaulmann et al., 2007), *S*-MBA would be quickly consumed to aminate the aromatic aldehyde leaving a lower concentration of *S*-MBA to aminate the dihydroxy ketones. In addition, the consumption of MBA leads to the reduction of pH and the acetophenone formation could inhibit the enzyme. The overall effect results in lower enzyme performance and aminodiol yield. Therefore, driving the TK reaction to completion is crucial for the subsequent reaction.

In an equimolar concentration reaction, the yield of 3-FBA reaction by D469T bearing mutants reached approximately 65 % conversion after 2 hours and this was suggested to be due to the product inhibition and HPA degradation by TK mutant (Payongsri et al., 2012) (and Chapter4, Table 4.4). Using a higher amount of HPA could, therefore, improve the yield by compensating the degraded HPA. The time course sampling of S385E/D469T/R520Q (Figure 6.3) showed that the increase in 3-DOPBA concentration, as well as the 3-FBA consumption rate, started to level out after 1 hour while HPA concentration still decreased. Addition of HPA at 1 hour after the reaction started may increase the chance of using the leftover HPA and consuming up 3-FBA.

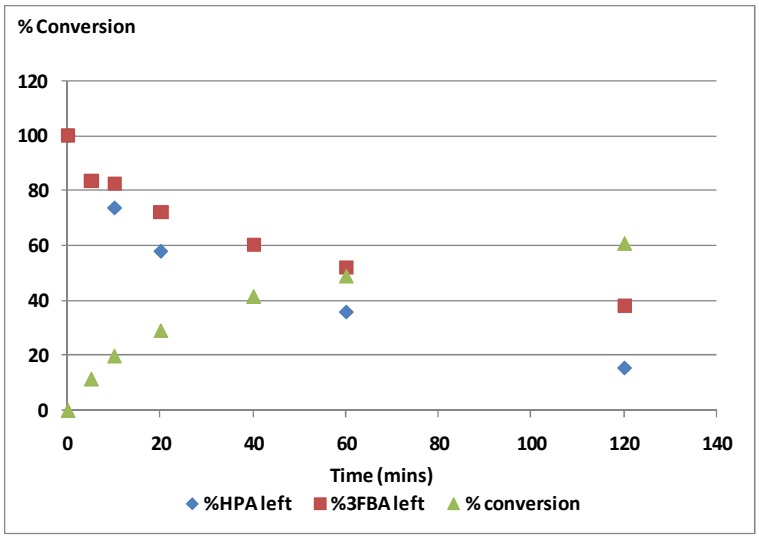


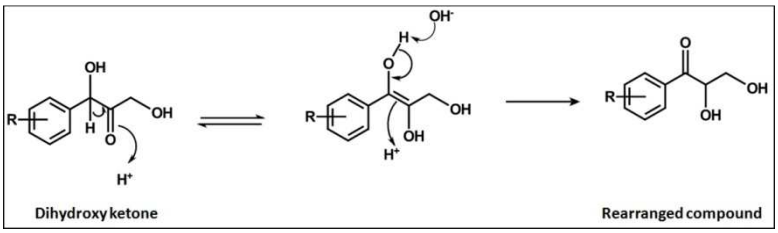
Figure 6.3 The reaction profile of equimolar3-FBA reaction catalysed by S385E/D469T/R520Q.

The 18-hour sample showed the concentration of 3-FBA was less than 3 mM left which illustrated that the fed-batch of HPA has improved the conversion yield to greater than 90 %. This percentage yield can still be achieved when the starting substrate concentrations were changed as long as 1.5 equivalent of HPA was added to the reaction at approximately 1 hour. However, using excess HPA at the beginning of the reaction does not improve the yield to the same degree.

Before the optimisation of the 4-FBA reaction at 3 hours after the 45-mM-equimolar reaction was started using S385Y/D469T/R520Q or S385T/D469T/R520Q, there was remaining 28 mM of 4-FBA, while the concentration of HPA was only 12 mM. This showed a greater degree of HPA degradation than the 3-FBA reaction and possibly due to the fact that the reactions of 4-FBA by both mutants were slower than the reaction of 3-FBA by S385E/D469T/R520Q (see Chapter5, Figure 5.8). Therefore, a greater degree of HPA degradation was observed. Adding 1.5 equivalent of HPA to the reaction at 3 hours has increased the 4-FBA consumption in both S385T/Y/D469T/R520Q to a similar level. The 18-hour samples showed that both mutants achieved greater than 70 % substrate consumption with 31 mM of 4-DOPBA.

3-DOPBA and 4-DOPBA from the TK reactions were used as 3X substrate solution and the 3-DOPBA and 4-DOPBA concentrations in the TAm reactions were kept constant at 10 mM. Therefore, the 3-FBA and 4-FBA reactions needed to be started with 35 mM and 45 mM, respectively. The concentrations of the dihydroxy ketone from both reactions were approximately 31 mM while the leftover 3-FBA and 4-FBA concentrations were less than 3

mM and 10 mM. Since some dihydroxy ketones underwent degradation, more aldehydes were consumed than the formation of the dihydroxy ketones. The major route of degradation was expected to be through the rearrangement at the 2- hydroxyl group as illustrated in Scheme 6.2(Galman et al., 2010). This was supported by a continuous increase in the formation of a compound that was highly absorbed at 250 nm.



Scheme 6.2 The degradation of the aromatic dihydroxy ketone through the rearrangement at the 2-hydroxyl group

(Galman et al., 2010).

6.3.2. MBA inhibition study

It was previously reported that MBA can have an inhibitory effect on CV2025 and other ω-TAms (Rios-Solis et al., 2011). Since the reaction with DPP was very slow compared with aldehydes (Kaulmann et al., 2007), the reaction rate could be increased by using higher substrate concentrations. In addition, using high substrate concentrations can also push the equilibrium forwards. Before screening with the aromatic dihydroxy ketones using MBA as a donor, it was therefore important to identify the maximum MBA concentration to be used without inhibiting the enzyme. Although one may speculate that MBA inhibition may occur at a different concentration when using different acceptor substrates, here, I identified if 40 mM MBA can inhibit CV2025 in 3-FBA/MBA reactions. The initial rates of the reaction at the concentration of 20 and 40 mM MBA were compared.

The equimolar reaction showed that the rate of 3-FBA consumption was 0.068 mM/min while the reaction with doubled MBA concentration had the initial rate of 0.069 mM/min. These initial rates were not significantly different and illustrated that 40 mM MBA can be used. If there is an inhibitory effect, it will have the same effect as at 20 mM. Although using higher substrate concentration does not seem to increase the initial rate, it still has the tendency to thermodynamically drive the reaction forwards. Therefore, 40 mM MBA was used through this study.

6.3.3. Screening transaminases for aromatic dihydroxy ketones

6.3.3.1. Screening with DPP

Prior to screening with 3-DOPBA and 4-DOPBA, the available ω-TAms were screened against DPP to find the best variants in accepting aromatic dihydroxy ketone in order to narrow down the variants for further studies. The initial screened illustrated that only the reaction of CV2025 showed a significant reduction of both substrates DPP and MBA with the highest acetophenone production which agreed with the previous report (Kaulmann et al., 2007). However, the CV2025 mutant Y153S and Y153F had marginal reduction in MBA concentration. As a result, CV2025 wild type was the best candidate for further screening. Slight deviations of the mass balance in all the reactions were observed possibly due to the evaporation of the reaction samples and acetophenone over 18 hours.

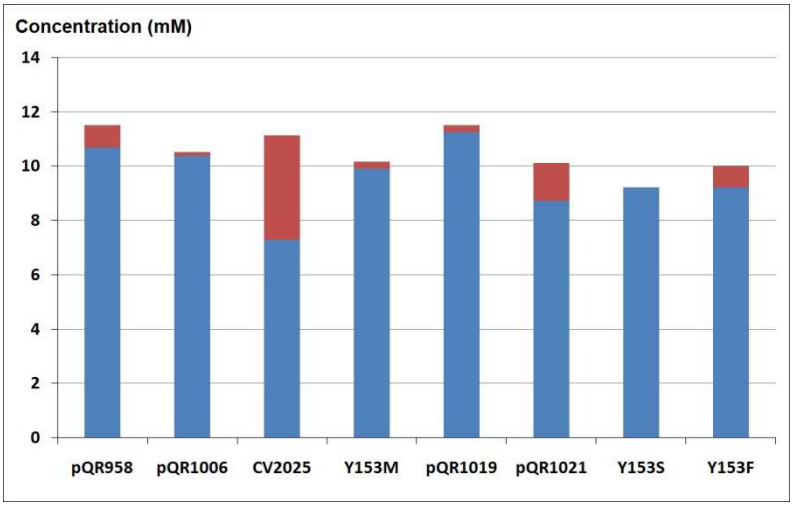


Figure 6.4 The acetophenone (■) and MBA (■) concentrations at 18 hours after the reaction between DPP and MBA was started.

The reaction was catalysed by different ω-TAms with the initial DPP and MBA concentrations of 10 mM.

6.3.3.2. Screening with 3-(1,3-Dihydroxy-2-oxopropyl)benzoic acid and 4-(1,3-Dihydroxy-2-oxopropyl)benzoic acid

Since the aromatic dihydroxy ketones are susceptible to degradation (Galman et al., 2010) especially when the temperature is elevated, my initial screening reactions were performed at 22 °C. Due to the unavailability of the amino alcohol standard, it was initially assumed that the production of acetophenone was the indication of the amino alcohol formation. However, my control reactions showed that clarified CV2025 lysate can catalyse the deamination of MBA and release acetophenone under the absence of external source

of amino acceptor. This deamination of MBA occurred in both Tris and HEPES buffers and up to 2 mM of acetophenone can be released within 3 hours when using 40 mM of MBA. This production of acetophenone and the consumption of MBA may be due to the amination of certain cellular metabolites such as pyruvate. When the dihydroxy ketones were added, at least 5 mM of acetophenone was produced while the less than 1 mM of the 3-DOPBA or 4-DOPBA was utilised and the presence of the leftover aldehydes from the TK reaction were less than 1 mM. Since acetophenone is produced even in the absence of any acceptor and the concentration is higher than the concentration of PLP alone, MBA was assumed to be used by CV2025 to aminate cellular metabolites.

In order to confirm that CV2025 can aminate the 3-DOPBA and 4-DOPBA, the chromatograms of all the reaction samples were overlayed and compared instead. In Figure 6.5, it can be seen that the peak area of the 3-DOPBA hardly changed. The aromatic amino alcohol is more polar than the dihydroxy ketones, so it was expected to be eluted much earlier. The superimposition the 2-minute sample chromatogram with the 16-hour sample showed that there was a small increase in the peak area with a retention time of 1.75 minutes which suggested that CV2025 may be able to accept the 3-DOPBA but that the reaction was very slow.

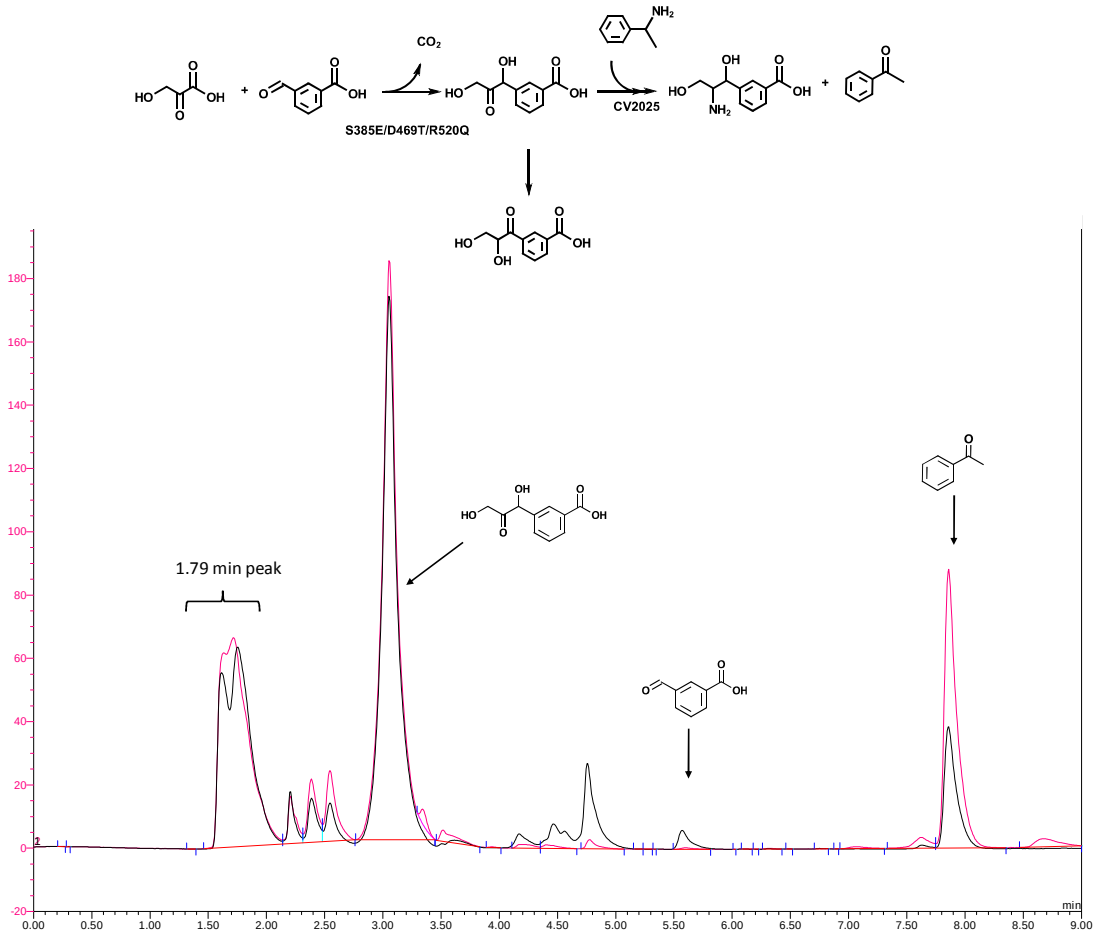


Figure 6.5 The overlapped chromatograms of the sample at 3 minutes (■) and 16 hours (■) of the 3-DOPBA amination reaction by CV2025.

The reaction was performed at 40 mM MBA, 10 mM 3-DOPBA, in 50 mM Tris buffer, pH 7.5, at 22 °C. The peaks of the compounds of interests were labelled with the structures corresponding to their retention times. The samples were analysed byACE5 C18 column and detected by 275 nm absorption.

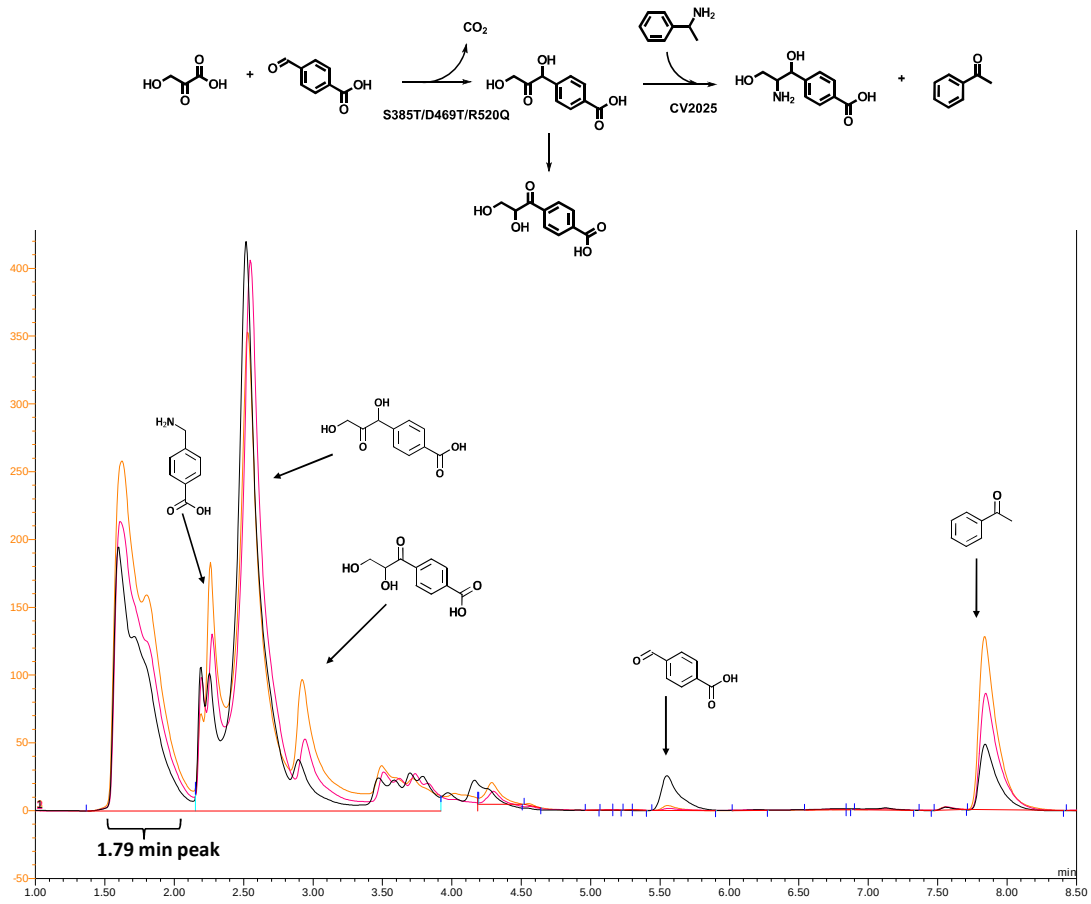


Figure 6.6 The overlapped chromatograms of the 4-DOPBA reaction samples at■2 minutes, ■6 hours, and ■26 hours.

The samples were detected by UV absorption at 275 nm. The reaction was performed at 40 mM MBA, 10 mM 4-DOPBA, in 50 mM Tris buffer, pH 7.5, at 22 °C . The peaks were labelled with the structures of the compounds corresponding to their retention times. The samples were analysed byACE5 C18 column and detected by 275 nm absorption.

The amination reaction of 4-DOPBA was also analysed by overlaying the chromatograms. In Figure 6.6, it can be seen that the 4-DOPBA peak in the 6-hour samples was slightly smaller than the 2-minute sample. A significant reduction was observed in the 26-hour sample. This may illustrate that CV2025 can catalyse the 4-DOPBA at a slightly better activity than 3-DOPBA but much slower than DPP. Alternatively, 4-DOPBA could undergo degradation over the period of time. This route was possible because the consumption of MBA leads to the reduction in the pH of the reaction which promotes the rearrangement. Interestingly, there was also an increase in the peak area at 1.6-1.8 minutes which was similar to what has been observed in the reaction of 3-DOPBA which may suggest that CV2025 may also accept the 4-DOPBA.

6.3.3.3. Optimising the reaction conditions for amination of aromatic dihydroxy ketone by CV2025

The slow amination reaction of aromatic dihydroxy ketones could have arisen from non-optimal conditions. The optimal pH for CV2025 was reported to be between 8.5-9 while the optimal temperature was 55-60 °C (Schell et al., 2009). However, my screening conditions were conducted at pH 7.5 and 22 °C which may slow down the overall TAm activity. Although higher temperature and pH may improve the performance of the TAm, lower amino alcohol yield may be expected because the degradation rates of the aromatic dihydroxy ketones also increase at higher temperature and pH. Consequently, the enzymatic reaction is competing with the thermal and chemical degradations. During the screening study, I observed that 4-FBA was aminated at a much higher rate than 3-FBA and the only difference between the two substrates is the position of the aldehyde relative to the carboxylic acid. It was then possible that 4-DOPBA would be more readily accepted by CV2025 than 3-DOPBA. The changes in the chromatograms also occurred more in the reaction of 4-DOPBA. Therefore, 4-DOPBA was chosen as a model substrate for condition optimisation. In this study, the amination of 4-DOPBA by CV2025 was performed under different combinations of pH values and temperatures in order to find the conditions that compromise the enzymatic activity and 4-DOPBA degradation. The differences in the amination rate of 3-FBA and 4-FBA will be further discussed in section 6.3.5.

In the pH 7.5 reactions, when the temperature was elevated from 22 °C to 30 °C and 37 °C, the acetophenone formation at 300 minutes increased from 7.35 to 8.15 and 10.5 mM respectively. This characteristic agreed with the previously determined optimal temperature where transaminase has higher activity as the temperature was elevated. In addition, the peak area at 1.79 minute also increased more rapidly at higher temperature. Subtracting the peak area of the sample at 300 minutes with 3-minute sample showed that the increase in this peak area doubled when the temperature was raised from 22°C to 37°C as showed in Figure 6.7a. However, the peak area of the 4-DOPBA at 22°C was similar to that of 30°C and 37°C until 6 hours. The major differences in the peak area of the 4-DOPBA among the three temperatures were observed in the 18-hour samples (Figure 6.7b) where greater than 50 % reduction of the 4-DOPBA was observed at 37°C compared with 33% and 32% reductions at 30°C and 22°C, respectively. The fact that all three temperatures had the same amount of 4-DOPBA left up until 6 hours while there were variations in the 1.79-

minute peaks suggested that the disappearance of the 4-DOPBA at 18 hours may not be due to the enzymatic reaction.

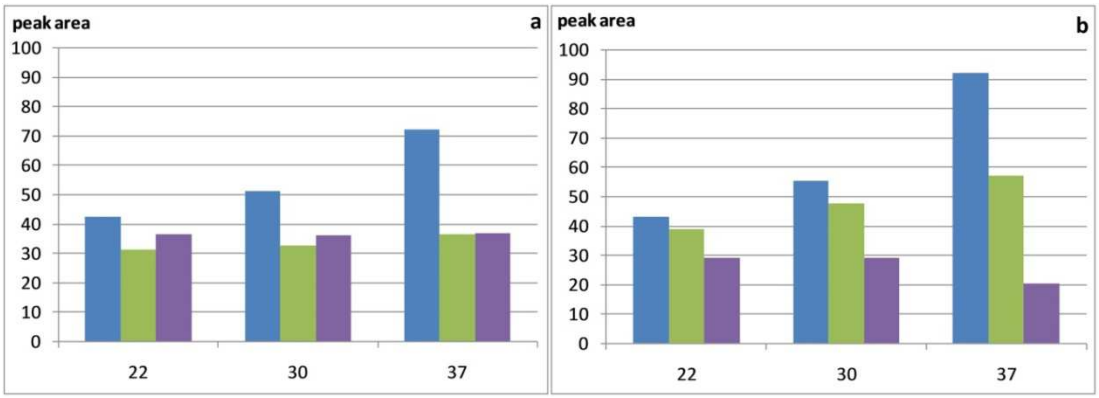


Figure 6.7 Comparison of the peak areas of different compounds at different temperatures, pH 7.5.
a) at 300 minutes and b) at 18 hours. ■ 1.79 min peak increased from 3 min, ■ rearranged compound (see Scheme 6.2) (250 nm detection), ■ remaining 4-DOPBA.

At pH 8.5 and 45 °C, both rearranged product and the 1.79-minute peak increased at higher proportion than the reaction at 37 °C with pH 7.5. However, the changes in 4-DOPBA concentrations up until 6 hours were similar at all temperatures. All the results so far suggested that it is unlikely that a further increase in the pH and temperature would promote the amination of 4-DOPBA by CV2025. However, the degradation of 4-DOPBA will become the major concern. Since none of the conditions appeared to increase the consumption of 4-DOPBA, it is likely that CV2025 cannot accept 4-DOPBA, and possibly 3-DOPBA.

In order to confirm whether the 1.79-minute peak is the amino alcohol, several control reactions were conducted. Changing the reaction buffer from Tris to HEPES still produced this peak in a similar proportion which suggested that the formation of this peak is independent to the buffer. When CV2025 clarified lysate was incubated with MBA without any external source of amino acceptor, this peak still became larger and this suggested that this peak cannot be the amino alcohol. In addition, when MBA was not added to the clarified lysate CV2025, the peak area still increased (Figure 6.8). However, this peak was not changed when usingonly his-tag purified CV2025 (Figure 6.9). All the evidence above suggested that cellular metabolites were the cause of the side reactions.

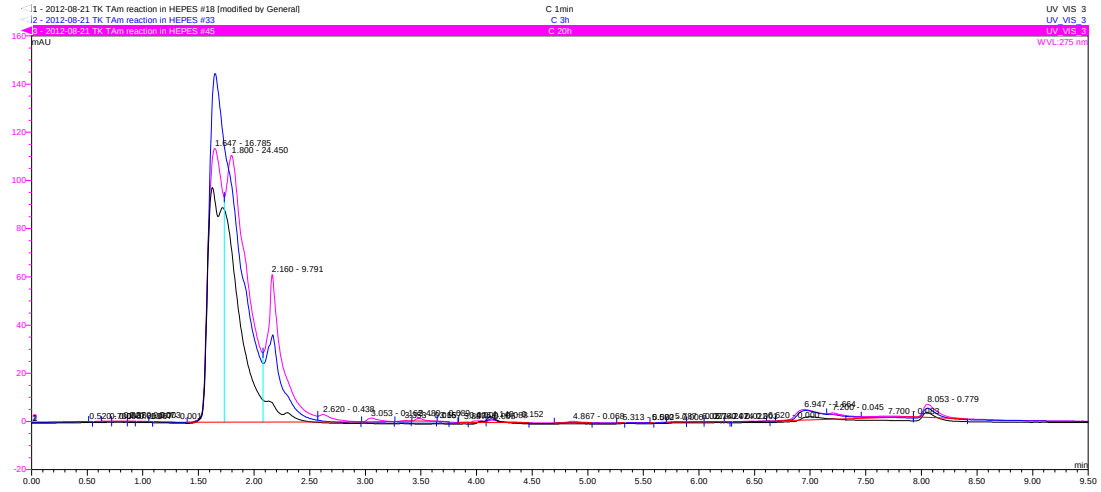


Figure 6.8 The overlapped chromatograms of CV2025 lysate when incubated with 50 mM Tris buffer, pH 7.5 with 0.4 mM PLP, 0.08 mM ThDP, and 0.3 $MgCl_2$ and at 30°C for (■) 1 minute, (■) 3 hours, and (■) 18 hours.

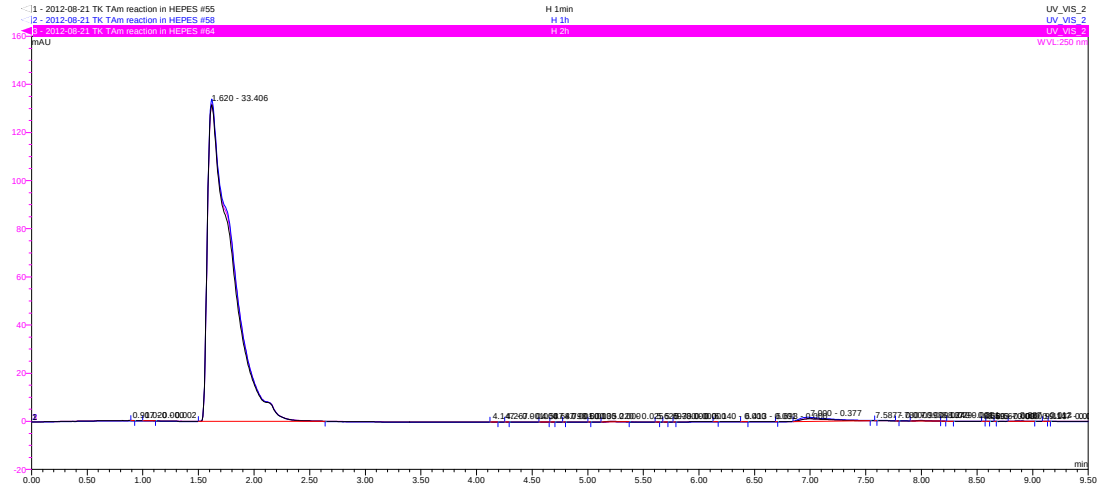


Figure 6.9 The overlapped chromatograms of purified CV2025 when incubated with 50 mM Tris buffer, pH 7.5 with 0.4 mM PLP, 0.08 mM ThDP, and 0.3 $MgCl_2$ and at 30°C for (■) 1 minute, (■) 1 hour, and (■) 2 hours.

6.3.4. Competitive reaction

All my evidence so far suggested that CV2025 accepted neither 3-DOPBA nor 4-DOPBA which led to a speculation whether both aromatic dihydroxy ketones could bind into the CV2025 active site in the first place. Since both 3-DOPBA and 4-DOPBA contain an aromatic ring and carboxylic group, one of these moieties may facilitate their binding as observed in the substrate spectrum of CV2025 (Kaulmann et al., 2007). In order to confirm whether 4-DOPBA could access into the active site of CV2025, a competitive reaction was performed. Here, the rates of 4-FBA amination were compared under the presence and absence of 4-DOPBA. If 4-DOPBA can access into the active site, it will slow down the rate of 4-FBA amination. In order to investigate the possibility that other components in the TK

reaction may inhibit CV2025 activity, one of the control reactions was set up by adding TK lysate and ThDP to the TAm reaction as showed in Table 6.1.

Table 6.1 The initial rate of 4-FBA amination under different conditions.

Reaction composition	Initial rate mM min^{-1}
4-FBA, 4-DOPBA, ThDP, TK	0.299
4-FBA, TK, ThDP	0.316
4-FBA	0.414

All the reactions were performed at 22°C , 40 mM MBA, 10 mM 4-FBA. The concentrations of ThDP in the first two reactions are 1.07 mM.

This competitive reaction showed that the rate of CV2025 was slowed down by 24% when TK lysate and ThDP were added alone. The presence of 4-DOPBA, however, did not significantly further reduce the rate of 4-FBA amination. These results suggested that the 4-DOPBA could not bind well in the active site of CV2025. Although ω-TAms may have dual substrate recognitions which allow them to bind to both aromatic substrates and alpha-keto acids (Cho et al., 2008; Shin and Kim, 2002), the presence of the carboxylic acid and aromatic ring of 4-DOPBA did not facilitate the amination at the carbonyl group of 4-DOPBA due to inaccessibility of the substrate itself. Since the competitive reaction between 3-FBA and 3-DOPBA was not performed, it cannot be ruled out that 3-DOPBA cannot access into the active site. The causes of the inability to catalyse 3-DOPBA and 4-DOPBA may be different due to their differences in the position of the carboxylic acid group. This was suggested by the kinetic study of 3-FBA and 4-FBA in the next section.

The analysis of the amino donors and acceptors of ω transaminase from *Vibrio fluvialis* JS17 suggested that the active site of ω transaminase consists of one large pocket and one small pocket (Shin and Kim, 2002) and CV2025 also show similar patterns and most dihydroxy ketones were poorly catalysed by CV2025 (Kaulmann et al., 2007). The initial rate of erythrulose and DPP aminations which were followed by the rate of acetophenone production suggested that erythrulose was aminated at approximately 4 times faster than DPP. This implied that the presence of the aromatic side chain appeared to significantly slow down the amination of the dihydroxy ketone in DPP (Kaulmann et al., 2007). Therefore, the inability of CV2025 to accept 3-DOPBA and 4-DOPBA was not surprising. Engineering to enlarge the active site is possible as illustrated in R-selective transaminase ATA-117 where the both binding pockets were enlarged to be able to accommodate prositagliptin (Savile et al., 2010).

6.3.5. Kinetics study for 3-FBA and 4-FBA

During the screening study of 3-DOPBA and 4-DOPBA, the crude TK reactions always have some 3-FBA and 4-FBA left in the reaction. It was noticed that these leftover aldehydes were not aminated at the same rate. This was then further looked into by assessing the specific activities towards both substrates at 10 mM aldehyde and 40 mM MBA under the standard conditions. The initial rate of 3-FBA amination was nearly 20 times slower than that of 4-FBA. Although it was previously illustrated that CV2025 preferred aromatic aldehydes (Kaulmann et al., 2007), such differences in the activities in the aromatic aldehydes have never been reported with any ω-TAms. In fact, ω-TAms was suggested to have dual substrate recognition through the hydrophobic interaction and the carboxylate group (Cho et al., 2008; Shin and Kim, 2002). It is rather impossible that the active site could differentiate 3-FBA and 4-FBA through hydrophobic interaction and the differences in the relative positions of the carboxylic acid group to the aldehyde must have been the cause of the difference in their activities, similar to my previous observation in TK (Payongsri et al., 2012). The kinetic parameters of the two aldehydes will illustrate the influences of the location of the carboxylic acid relative to the carbonyl group on the enzymatic activity. This could give an insight into the active site of CV2025 and the substrate recognition which could benefit the active site engineering.

The relationship between the initial rate of 3-FBA amination and 3-FBA concentration in Figure 6.10 showed that 3-FBA strongly inhibited CV2025 activity and an inhibitory concentration was observed at 3 mM of 3-FBA. Monitoring the reaction below 3 mM was found to be inaccurate due to the detection limit. Consequencely, the concentration at which 3-FBA started to inhibit the enzyme was not determined. The specific activity at 15 mM was found to be 0.076 µmol/mg/min while the maximum specific activity was 0.25 µmol/mg/min at 3 mM.

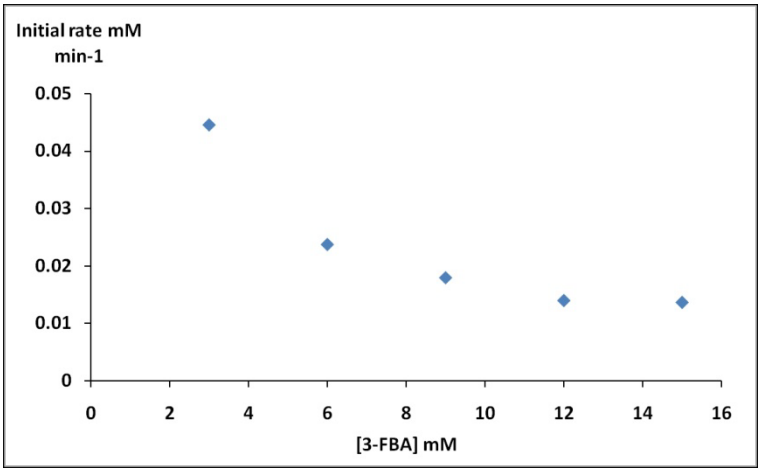


Figure 6.10 The relationship between the 3-FBA concentration and the initial rate of 3-FBA amination by CV2025.

All substrates were prepared in 50 mM Tris buffer, pH 7.5. The MBA concentration was kept constant at 50 mM and the enzyme concentration in the reaction was 0.18 mg/min.

In contrast, 4-FBA was not found to inhibit the reaction at 15 mM. However, the results did not fit to a simple Michaelis-Menten curve either but rather a sigmoidal curve seemed to fit the data instead (Figure 6.11). This sigmoidal curve suggested that there could be allosteric control although both MBA and 4-FBA are non-natural substrates for CV2025. The reaction at very low 4-FBA concentration was found to be very fast and the full conversion was achieved instantly after the addition of MBA. Therefore, the initial rate at the concentration lower than 3 mM was not studied. The maximum specific activity was 5.15 µmol/mg/min which makes the k_{cat} become 4.4 s^{-1}. This specific activity is much higher than the activity of the best variant TK for 4-FBA which makes the true one-pot synthesis rather improbable.

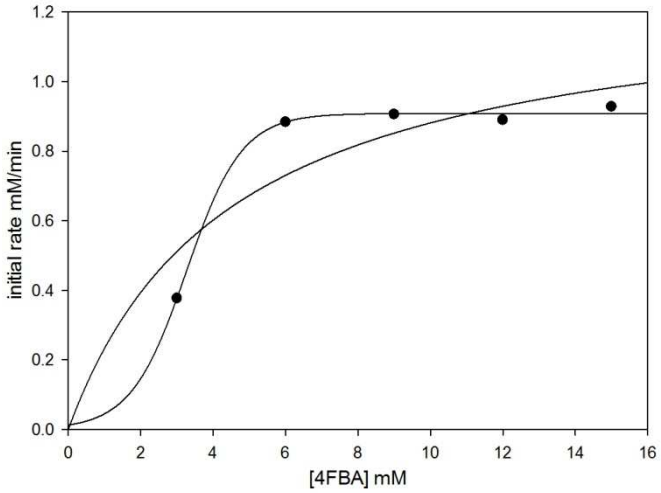


Figure 6.11 The relationship between the 4-FBA concentration and the initial rate of 4-FBA amination.
The Michaelis-Menten and sigmoid curves were fitted into the data for comparison. All substrates were prepared in 50 mM Tris buffer, pH 7.5. MBA concentration was kept constant at 50 mM and the enzyme concentration in the reaction was 0.18 mg/ml.

CV2025 has very high affinity towards both 3-FBA and 4-FBA and such affinity could arise from the electrostatic interaction between the enzyme and substrates. The residue that could interact with the carboxylate group is the highly conserver R416 of which the guanidinium group was located 10 Å away from the nitrogen atom of PLP. This distance is long enough to accommodate 4-FBA molecule and place the aldehyde group close to the reactive centre. The fact that 4-FBA has such a high affinity towards CV2025, it may bind to the active site so rapidly that 4-DOPBA could not compete with (fast k_{on}, slow k_{off}). However, this does mean that 4-DOPBA did not employ electrostatic interaction in a similar way to 4-FBA. Alternatively, very high steric hindrance elsewhere in the active site may destabilise the binding (fast k_{off}).

6.3.6. Active site structure and possible reasons for slow reaction with aromatic dihydroxy ketones and 3-FBA

If the carboxylate group of 3-FBA and 4-FBA were bound to the same positively charged amino acid at the active site of CV2025, the positions of the aldehyde groups in 3-FBA and 4-FBA were only 1 bond-length difference. Yet, this created such difference in their kinetic properties. Comparing the position of the carbonyl group relative to the carboxylic acid in 4-FBA and 4-DOPBA, it can be seen that the carbonyl group in 4-DOPBA is 1 C-C bond further away from the reactive centre of PLP. In addition, the carbonyl group could orient in a direction that did not favour the reaction to process. If this is true, the situation in 3-DOPBA could be worse because the distance between the carbonyl group and the PLP

reactive centre would be even further. The fact that DPP could be accepted by CV2025 also supports this hypothesis. The lack of the carboxylic acid group on DPPwould allow this molecule to move more freely within the active site and increase the chance of amination at the carbonyl group of DPP.

However, our competitive reaction illustrated that 4-DOPBA, and possibly 3-DOPBA, may not be able to access into the active site of CV2025. Therefore, there could be some other steric hindrance upon the binding of both compounds. In fact, the crystal structure of CV2025 (PDB ID: 4A6T) (Humble et al., 2012) showed that the active site was a narrow channel mainly consisting of aromatic amino acids as showed in Figure 6.12a and b. The aromatic rings of two amino acids Y153 and W60 were sandwiched and create a small channel which was suggested to be important in the stereoselective control of ω-TAms (Shin and Kim, 2002). This was confirmed by the ability to enhance the stereoselectivity by a single mutation at W60 (Humble et al., 2012). The width of this channel due to the sandwich packing between Y153 and W60 was 7.56 Å. A 90°-rotation view showed that A231 and F88 constricted the active site which was due to the fact that their side chains were protruding perpendicularly to the planes of T153 and W60. However, the distance between A231 and F88 was approximately 8.64 Å which was large enough to allow erythrulose and other molecules with dihydroxy ketone moiety to reach PLP, including DPP. These 4 amino acids could be the target for site directed mutagenesis study to improve the enzymatic activity towards aromatic dihydroxy ketones. However, a few residues may need to be mutated simultaneously. In addition, the kinetic study of 3-FBA and 4-FBA illustrated that the position of the carbonyl group is also crucial. Since TK extends 1 additional bond from the carbonyl group, repositioning the whole aromatic dihydroxy ketone may need to be considered to allow the carbonyl group to be places at close proximity to the reactive centre.

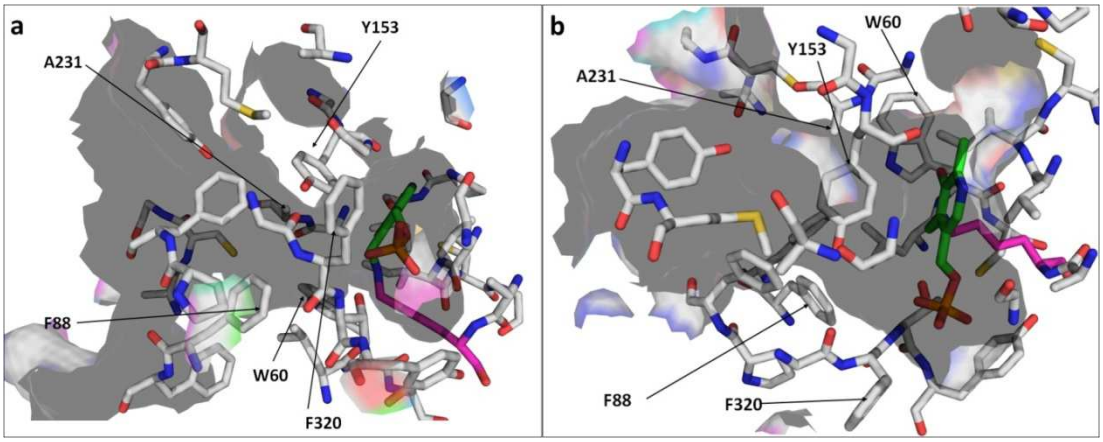


Figure 6.12 The active site channel of CV2025 (PDB ID: 4A6T). The channel was illustrated as the surface of amino acids constituting the active site.

a) the side view image and b) the 90°-rotated image of a). The PLP was showed in green stick and catalytic Lys288 (as showed in Figure 6.1) was showed in magenta stick.

6.4. Conclusion

The enzymatic synthesis of 3-DOPBA and 4-DOPBA has been optimised by addition of HPA at a certain time after the reaction started. The conversion yield can be up to 90 % and 70% yields for 3-FBA and 4-FBA respectively. However, CV2025 did not appear to be able to aminate either 3-DOPBA or 4-DOPBA to form aromatic amino alcohols. The competitive reaction on 4-DOPBA suggested that this was due to inaccessibility to the active site of 4-DOPBA itself. Some residues were suggested to be the target for mutagenesis.

7. Conclusion and future research recommendation

The novel synthetic pathway of aminoalcohol by transketolase and transaminase was proposed to be a more effective and efficient synthetic pathway than the traditional organic synthetic routes. In the past, glycolaldehyde and propionaldehyde were extensively studied in the above enzymatic system. When using different aromatic aldehydes, derivatives of chloramphenicol can also be synthesised. However, only benzaldehyde has ever been used as a substrate for certain transketolase mutants and the yield of 1,3-dihydroxy-1-phenylpropane-2-one was far too low to be used as a substrate for transaminase reaction and only the synthetic compound was ever used. Since the yield and activity of transketolase with aromatic aldehyde was the bottleneck, it can limit the application of this novel pathway to just aliphatic aldehydes.

In order to expand the application of this novel pathway, the overall goal of this research was to construct a small library of transketolase mutants with higher activities towards different aromatic aldehydes in order to synthesise novel aromatic dihydroxy ketones. The aromatic dihydroxy ketones were then fed into the transaminase reaction to produce aromatic aminoalcohols which can be used as a building block for chloramphenicol derivatives. The stability of the accumulation of mutation was also considered because accumulation of mutations usually destabilise the protein and lead to aggregate formation and lower enzyme titre. Therefore, the desired parental mutant must have high activity as well as stability.

During this research, the quantification of the enzyme concentration in the lysate (densitometry) was revisited and redeveloped together with the reaction monitoring procedure in order to achieve reliable and repeatable specific activity of the enzyme. The degradation of hydroxypyruvate by transketolase mutants and a number of side reactions were observed during this research. They were speculated to compete with the ketol transfer reaction which also resulted in lower final yields if the specific activity of the reaction is low. For a slow to moderate reaction rate, hydroxypyruvate degradation has a more pronounced effect on the yield. As a result, the improved yields can be used as an indicator for enhanced activity. The final yield could still be further improved by the fed-batch regime of hydroxypyruvate. These side reactions required a new screening method that is more accurate and specific than the colorimetric assay and HPLC was inevitably developed as a tool although it has lower throughput. This methodology was shown to be compatible for screening whole cell biocatalyst.

In the previous studies (Hibbert et al., 2008, 2007), several singletransketolase mutants were found to have higher activities towards glycolaldehyde and propionaldehyde. In the attempt to further improve their activities, two transketolase libraries were created. Random combination of active mutants was showed to be ineffective to produce highly active mutants but it led to unstable enzymes. In contrast, considering co-evolved residues and finding the right pair to be simultaneously replaced has showed to stabilise the enzyme as well as maintaining the activity.H26 and H461 were showed to be inappropriate target because their mutations drastically reduced the enzyme stability and the overall enzymatic activities for both substrates.

Prior to constructing the new transketolase library for aromatic aldehydes, the available transketolase crystal structures were analysed and a rational design approach was employed to find the mutants that accepted aromatic aldehyde. The use of natural substrate analogues, namely 3-formylbenzoic acid (3-FBA) and 4-formylbenzoic acid (4-FBA), has showed that D469T was the best mutant and it should be the template for further protein engineering. The enzyme-substrate affinity and the substrate orientation within the active sites were the two key factors controlling the bioconversion rate of the aromatic aldehydes. Computational docking of both substrates into D469T and D469T/R520Q suggested that the carboxylate group on both substrates interacts with the positively charged residues that bind to the phosphate group of natural substrates.

The study of a pair of co-evolved residues in Chapter 3 suggested that D469T/R520Q is more stable than D469T. Site-saturation mutagenesis at the residues within the first shell produced several triple mutants with higher activities and yields towards 3-FBA, 4-FBA, and 3-hydroxybenzaldehyde. The kinetic study of these mutants illustrated that the k_{cat} of each mutant differ from one aldehyde to another which implied that these aromatic aldehyde with different substituents bind to the active site in different orientations.At this stage, the overall goal of the research was reached.

The optimised 3-FBA and 4-FBA reactions were fed into the transaminase reaction in order to synthesise novel aromatic dihydroxy ketones in Chapter 6. Unfortunately, none of the available omega transaminases including CV2025 appeared to accept any newly synthesised aromatic dihydroxy ketones. The bottleneck for the transketolase-transaminase pathway was identified as the inability of transaminase to accept the aromatic dihydroxy ketone with ring substituents.

The current outcomes have showed the potential and also new obstacles of the transketolase-transaminase reaction and they have to be carefully evaluated for the research direction.

7.1. Future research recommendation for transketolase

David Steadman (Chemistry Department) and I have explored several additional aromatic aldehydes with 3-substituents. However, the 4-substituents have not been fully researched and very few mutants have rather moderate activities for 4-FBA. For an immediate plan, aromatic aldehydes with 4-substituents could be further explored and the S385X/D469T/R520Q library can still be a good starting point. Since most of the mutants have relatively low k_{cat} of the 4-FBA, it is still possible that additional mutations can further improve the activity of the enzyme. In addition, other D469 variants could be investigated even the D469Y mutant. Mutation at this residue has been showed to alter the stereoselectivity as well as improving the activity for propionaldehyde before (Smith et al., 2008) but several other residues may need to be simultaneously mutated in order to allow D469Y to accept aromatic aldehydes. Besides the research in engineering transketolase, the enzymatic synthesis 3-(1,3-Dihydroxy-2-oxopropyl)benzoic acid (3-DOPBA) and 4-(1,3-Dihydroxy-2-oxopropyl)benzoic acid (4-DOPBA) may be used as a substrate in traditional chemical synthesis to produce two different chloramphenicol analogues and tested their antibacterial properties.

It should be noted that most of the screening methods only quantify the product concentration while the stereoselectivity of the mutants cannot be immediately determined. It should be noted that a parallel study by David Steadman illustrated that the absolute stereoselectivity depends on the substituent and the highly active mutants appear to have *S* selective for most aldehydes with the 3- substituent. A different strategy for designing transketolase library will need to be considered to alter the stereoselectivity. Amino acid addition or deletion may be considered to alter the position of the acid/base catalysts relative to the aldehyde. The availability of the crystal structure of the mutants especially the substrate-bound structures would significantly facilitate the next library design and computational docking of other substrates.

Most of the neutral-charged aromatic aldehydes are poorly soluble in aqueous solution while the K_M values for these aromatic aldehydes are likely to be large. If the apparent aldehyde concentration is higher, possibly with the aid of co-solvent, the

bioconversion rate could be further enhanced. However, this strategy requires transketolase and transaminase mutants that well tolerate organic solvents. Certain mutants have been studied and shown to meet the requirement and they can be cooperated into the highly active mutants for this purpose. The bioconversion rate is a critical parameter because it has a direct impact on the productivity and the overall production cost which determines the economic viability of the process.

7.2. Future research recommendation for transketolase-transaminase pathway for the synthesis of aromatic aminoalcohols

The experiment in Chapter 6 illustrated that the major bottleneck of the transketolase-transaminase pathway is the transaminase reaction. This is because none of the available transaminases appeared to accept aromatic dihydroxy ketones with ring substituent and most of the omega transaminases are also *S* selective. However, some R-selective transaminases are commercially available and they should be tested with 3-DOPBA and 4-DOPBA to assess the process feasibility and other challenges.

The equilibrium is still one of the major issues for most transaminase reactions especially when cheaper amino donors such as isopropylamine or amino acids are to be used and this issue makes the overall yield very low. To overcome this issue, an engineering strategy such as product removal is required to overcome this issue. However, the final yield, production rate, the overall production cost and its commercial value has to be evaluated since the final product is still cheaply available.

7.3. Overall evaluation and recommendation

The research at this stage suggested that the synthesis of chloramphenicol and its analogues by transketolase-transaminase cascade is unlikely to be feasible. For economic evaluation, the cost of starting materials from both reaction especially hydroxypyruvic acid and (S)-methylbenzylamine is still higher than the selling price of chloramphenicol. To minimise the reagent cost, an alternative substrate, isopropylamine, could be used. In addition, the yield of the process would be very low due to the equilibrium position of the transaminase reaction and this could make the final product uneconomically viable. Using excess isopropylamine and the removal of acetone which is the product of isopropylamine deaminationby applying negative pressure, the reaction could be driven to completion. Nonetheless, this research has its own merit. This thesis unravelled the key mechanism of transketolase for non-natural substrates which can give further insight into the

understanding of the substrate specificity of the enzymes within the same family. Other enzyme within the same family may be considered such as pyruvate decarboxylase which already gives *R* selective for the ligation between pyruvate and benzaldehyde (Sprenger and Pohl, 1999).

8. References

Arnold, F.H., 1998. Design by Directed Evolution. Accounts of Chemical Research 31, 125–131.

Arnold, F.H., Wintrode, P.L., Miyazaki, K., Gershenson, A., 2001. How enzymes adapt: lessons from directed evolution. Trends in biochemical sciences 26, 100–6.

Asztalos, P., Parthier, C., Golbik, R., Kleinschmidt, M., Hübner, G., Weiss, M.S., Friedemann, R., Wille, G., Tittmann, K., 2007. Strain and near attack conformers in enzymic thiamin catalysis: X-ray crystallographic snapshots of bacterial transketolase in covalent complex with donor ketoses xylulose 5-phosphate and fructose 6-phosphate, and in noncovalent complex with acceptor aldo. Biochemistry 46, 12037–52.

Aucamp, J.P., Martinez-Torres, R.J., Hibbert, E.G., Dalby, P.A., 2008. A microplate-based evaluation of complex denaturation pathways: structural stability of Escherichia coli transketolase. Biotechnology and bioengineering 99, 1303–10.

Behrens, G.A., Hummel, A., Padhi, S.K., Schätzle, S., Bornscheuer, U.T., 2011. Discovery and Protein Engineering of Biocatalysts for Organic Synthesis. Advanced Synthesis & Catalysis 353, 2191–2215.

Bloom, J.D., Labthavikul, S.T., Otey, C.R., Arnold, F.H., 2006. Protein stability promotes evolvability. Proceedings of the National Academy of Sciences of the United States of America 103, 5869–74.

Bolon, D.N., Voigt, C. a, Mayo, S.L., 2002. De novo design of biocatalysts. Current opinion in chemical biology 6, 125–9.

Bolte, J., Demuynck, C., Samaki, H., 1987. Utilization of enzymes in organic chemistry: Transketolase catalyzed synthesis of ketoses. Tetrahedron Letters 28, 5525–5528.

Bornscheuer, U.T., Pohl, M., 2001. Improved biocatalysts by directed evolution and rational protein design. Current opinion in chemical biology 5, 137–43.

Bougioukou, D., Kille, S., Taglieber, A., Reetz, M., 2009. Directed Evolution of an Enantioselective Enoate-Reductase: Testing the Utility of Iterative Saturation Mutagenesis. Advanced Synthesis & Catalysis 351, 3287–3305.

Breslow, R., 1957. RAPID DEUTERIUM EXCHANGE IN THIAZOLIUM SALTS 1. Journal of the American Chemical Society 79, 1762–1763.

Breslow, R., 1958. On the Mechanism of Thiamine Action. IV. 1 Evidence from Studies on Model Systems. Journal of the American Chemical Society 80, 3719–3726.

Breuer, M., Ditrich, K., Habicher, T., Hauer, B., Kesseler, M., Stürmer, R., Zelinski, T., 2004. Industrial methods for the production of optically active intermediates. Angewandte Chemie (International ed. in English) 43, 788–824.

Breuer, M., Hauer, B., 2003. Carbon–carbon coupling in biotransformation. Current Opinion in Biotechnology 14, 570–576.

Brown, N.G., Pennington, J.M., Huang, W., Ayvaz, T., Palzkill, T., 2010. Multiple global suppressors of protein stability defects facilitate the evolution of extended-spectrum TEM β-lactamases. Journal of molecular biology 404, 832–46.

Cadwell, R.C., Joyce, G.F., 1992. Randomization of genes by PCR mutagenesis. Genome Research 2, 28–33.

Calloni, G., Zoffoli, S., Stefani, M., Dobson, C.M., Chiti, F., 2005. Investigating the effects of mutations on protein aggregation in the cell. The Journal of biological chemistry 280, 10607–13.

Cázares, A., Galman, J.L., Crago, L.G., Smith, M.E.B., Strafford, J., Ríos-Solís, L., Lye, G.J., Dalby, P.A., Hailes, H.C., 2010. Non-alpha-hydroxylated aldehydes with evolved transketolase enzymes. Organic & biomolecular chemistry 8, 1301–9.

Ceccarelli, E.A., Carrillo, N., Roveri, O.A., 2008. Efficiency function for comparing catalytic competence. Trends in biotechnology 26, 117–8.

Cedrone, F., Ménez, A, Quéméneur, E., 2000. Tailoring new enzyme functions by rational redesign. Current opinion in structural biology 10, 405–10.

Chen, B.H., Sayar, A., Kaulmann, U., Dalby, P.A., Ward, J.M., Woodley, J.M., 2006. Reaction modelling and simulation to assess the integrated use of transketolase and ω-transaminase for the synthesis of an aminotriol. Biocatalysis and Biotransformation 24, 449–457.

Chen, Z., Zhao, H., 2005. Rapid creation of a novel protein function by in vitro coevolution. Journal of molecular biology 348, 1273–82.

Chica, R.A., Doucet, N., Pelletier, J.N., 2005. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Current opinion in biotechnology 16, 378–84.

Cho, B.K., Park, H.Y., Seo, J.H., Kim, J., Kang, T.J., Lee, B.S., Kim, B.G., 2008. Redesigning the substrate specificity of omega-aminotransferase for the kinetic resolution of aliphatic chiral amines. Biotechnology and bioengineering 99, 275–84.

Controulis, J., Rebstock, M.C., Crook, H.M., 1949. Chloramphenicol (chloromycetin). Journal of the American Chemical Society 71, 2463–2468.

Corbett, M.D., Corbett, B.R., 1986. Effect of ring substituents on the transketolase-catalyzed conversion of nitroso aromatics to hydroxamic acids. Biochemical pharmacology 35, 3613–21.

Costelloe, S.J., Ward, J.M., Dalby, P.A., 2008. Evolutionary analysis of the TPP-dependent enzyme family. Journal of molecular evolution 66, 36–49.

Dalby, P.A., 2003. Optimising enzyme function by directed evolution. Current Opinion in Structural Biology 13, 500–505.

Dalby, P.A., 2007. Engineering enzymes for biocatalysis. Recent patents on biotechnology 1, 1–9.

Dalby, P.A., 2011. Strategy and success for the directed evolution of enzymes. Current opinion in structural biology 21, 473–80.

Dalby, P.A., Aucamp, J.P., George, R., Martinez-Torres, R.J., 2007. Structural stability of an enzyme biocatalyst. Biochemical Society transactions 35, 1606–9.

Delihas, N., Forst, S., 2001. MicF: an antisense RNA gene involved in response of Escherichia coli to global stress factors. Journal of molecular biology 313, 1–12.

Demuynck, C., Bolte, J., Hecquet, L., Dalmas, V., 1991. Enzyme-catalyzed synthesis of carbohydrates: synthetic potential of transketolase. Tetrahedron Letters 32, 5085–5088.

Demuynck, C., Bolte, J., Hecquet, L., Samaki, H., 1990. Enzymes as reagents in organic chemistry: transketolase-catalysed synthesis of d-[1,2-13C2]xylulose. Carbohydrate Research 206, 79–85.

Drummond, D.A., Iverson, B.L., Georgiou, G., Arnold, F.H., 2005. Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. Journal of molecular biology 350, 806–16.

Eisenthal, R., Danson, M.J., Hough, D.W., 2007. Catalytic efficiency and kcat/KM: a useful comparator? Trends in biotechnology 25, 247–9.

Enders, D., Narine, A.A., 2008. Lessons from nature: biomimetic organocatalytic carbon-carbon bond formations. The Journal of organic chemistry 73, 7857–70.

Enders, D., Voith, M., Lenzen, A., 2005. The dihydroxyacetone unit--a versatile C(3) building block in organic synthesis. Angewandte Chemie (International ed. in English) 44, 1304–25.

Esakova, O.A., Meshalkina, L.E., Kochetov, G.A., 2005. Effects of transketolase cofactors on its conformation and stability. Life sciences 78, 8–13.

Esakova, O.A., Meshalkina, L.E., Kochetov, G.A., Golbik, R., 2009. Halogenated pyruvate derivatives as substrates of transketolase from Saccharomyces cerevisiae. Biochemistry (Moscow) 74, 1234–1238.

Fessner, W.D., 1998. Enzyme mediated C-C bond formation. Current opinion in chemical biology 2, 85–97.

Fiedler, E., Golbik, R., Schneider, G., Tittmann, K., Neef, H., König, S., Hübner, G., 2001. Examination of donor substrate conversion in yeast transketolase. The Journal of biological chemistry 276, 16051–8.

Fox, R.J., Clay, M.D., 2009. Catalytic effectiveness, a measure of enzyme proficiency for industrial applications. Trends in biotechnology 27, 137–40.

Frank, R.A.W., Titman, C.M., Pratap, J.V., Luisi, B.F., Perham, R.N., 2004. A molecular switch and proton wire synchronize the active sites in thiamine enzymes. Science 306, 872–6.

French, C., Ward, J.M., 1995. Improved production and stability ofE. coli recombinants expressing transketolase for large scale biotransformation. Biotechnology Letters17, 247–252.

Fry, K., Ingraham, L.L., Westheimer, F.H., 1957. The Thiamin-Pyruvate Reaction. Journal of the American Chemical Society 79, 5225–5227.

Galman, J.L., Steadman, D., Bacon, S., Morris, P., Smith, M.E.B., Ward, J.M., Dalby, P.A., Hailes, H.C., 2010. α,α'-Dihydroxyketone formation using aromatic and heteroaromatic aldehydes with evolved transketolase enzymes.Chemical communications (Cambridge, England) 46, 7608–10.

Galman, J.L., Steadman, D., Haigh, L.D., Hailes, H.C., 2012. Investigating the reaction mechanism and organocatalytic synthesis of α,α'-dihydroxy ketones. Organic & biomolecular chemistry 10, 2621–8.

Glasner, M.E., Gerlt, J. a, Babbitt, P.C., 2006. Evolution of enzyme superfamilies.Current opinion in chemical biology 10, 492–7.

Göbel, U., Sander, C., Schneider, R., Valencia, A., 1994. Correlated mutations and residue contacts in proteins. Proteins 18, 309–17.

Hailes, H., Dalby, P.A., Lye, G., Ward, J., 2009. Biocatalytic approaches to ketodiols and aminodiols. Chimica Oggi Chemistry Today 27, 28–31.

Hailes, H.C., Dalby, P.A., Lye, G., Baganz, F., Micheletti, M., Szita, N., Ward, J., 2010. alpha,alpha ’-Dihydroxy Ketones and 2-Amino-1,3-diols: Synthetic and Process Strategies Using Biocatalysts. Current Organic Chemistry 14, 1883–1893.

Hailes, Helen C., Dalby, P.A., Lye, G.J., Ward, J.M., 2010. ChemInform Abstract: Biocatalytic Approaches to Ketodiols and Aminodiols. ChemInform 41, no–no.

Hawkins, C.F., Borges, A., Perham, R.N., 1989. A common structural motif in thiamin pyrophosphate-binding enzymes. FEBS letters 255, 77–82.

Hayes, R.J., Bentzien, J., Ary, M.L., Hwang, M.Y., Jacinto, J.M., Vielmetter, J., Kundu, A., Dahiyat, B.I., 2002. Combining computational and experimental screening for rapid optimization of protein properties. Proceedings of the National Academy of Sciences of the United States of America. 99, 15926–31.

Hecquet, L., Bolte, J., Demuynck, C., 1994. Chemoenzymatic synthesis of 6-deoxy-D-fructose and 6-deoxy-L-sorbose using transketolase. Tetrahedron 50, 8677–8684.

Hecquet, L., Bolte, J., Demuynck, C., 1996. Enzymatic synthesis of “natural-labeled” 6-deoxy-L-sorbose precursor of an important food flavor. Tetrahedron 52, 8223–8232.

Hibbert, E.G., Senussi, T., Costelloe, S.J., Lei, W., Smith, M.E.B., Ward, J.M., Hailes, H.C., Dalby, P.A., 2007. Directed evolution of transketolase activity on non-phosphorylated substrates. Journal of biotechnology 131, 425–32.

Hibbert, E.G., Senussi, T., Smith, M.E.B., Costelloe, S.J., Ward, J.M., Hailes, H.C., Dalby, P.A., 2008. Directed evolution of transketolase substrate specificity towards an aliphatic aldehyde.Journal of biotechnology 134, 240–5.

Hobbs, G.R., Lilly, M.D., Turner, N.J., Ward, J.M., Willets, A.J., Woodley, J.M., 1993. Enzyme-catalysed carbon?carbon bond formation: use of transketolase from Escherichia coli. Journal of the Chemical Society, Perkin Transactions 1 165.

Höhne, M., Bornscheuer, U.T., 2009. Biocatalytic Routes to Optically Active Amines. ChemCatChem 1, 42–51.

Hübner, G., Tittmann, K., Killenberg-Jabs, M., Schäffner, J., Spinka, M., Neef, H., Kern, D., Kern, G., Schneider, G., Wikner, C., Ghisla, S., 1998. Activation of thiamin diphosphate in enzymes.Biochimica et biophysica acta 1385, 221–8.

Humble, Maria Svedendahl, Cassimjee, K.E., Abedi, V., Federsel, H.-J., Berglund, P., 2012. Key Amino Acid Residues for Reversed or Improved Enantiospecificity of an ω-Transaminase. ChemCatChem 4, 1167–1172.

Humble, Maria S, Cassimjee, K.E., Håkansson, M., Kimbung, Y.R., Walse, B., Abedi, V., Federsel, H.-J., Berglund, P., Logan, D.T., 2012. Crystal structures of the Chromobacterium violaceumω-transaminase reveal major structural rearrangements upon binding of coenzyme PLP. The FEBS journal 279, 779–92.

Humphrey, A.J., Parsons, S.F., Smith, M.E.., Turner, N.J., 2000. Synthesis of a novel N-hydroxypyrrolidine using enzyme catalysed asymmetric carbon–carbon bond synthesis. Tetrahedron Letters 41, 4481–4485.

Humphrey, A.J., Turner, N.J., McCague, R., Taylor, S.J.C., 1995. Synthesis of enantiomerically pure ?-hydroxyaldehydes from the corresponding ?-hydroxycarboxylic acids: novel substrates for Escherichia coli transketolase. Journal of the Chemical Society, Chemical Communications 2475.

Ingram, C.U., Bommer, M., Smith, M.E.B., Dalby, P.A., Ward, J.M., Hailes, H.C., Lye, G.J., 2007. One-pot synthesis of amino-alcohols using a de-novo transketolase and beta-alanine: pyruvate transaminase pathway in Escherichia coli. Biotechnology and bioengineering 96, 559–69.

Ishige, T., Honda, K., Shimizu, S., 2005. Whole organism biocatalysis. Current opinion in chemical biology 9, 174–80.

Jahromi, R.R.F., Morris, P., Martinez-Torres, R.J., Dalby, P.A., 2011. Structural stability of E. coli transketolase to temperature and pH denaturation. Journal of biotechnology 155, 209–16.

Jover, J., Bosque, R., Sales, J., 2008. QSPR Prediction of pKa for Benzoic Acids in Different Solvents. QSAR & Combinatorial Science 27, 563–581.

Kaulmann, U., Smithies, K., Smith, M.E.B., Hailes, H.C., Ward, J.M., 2007. Substrate spectrum of ω-transaminase from Chromobacterium violaceum DSM30191 and its potential for biocatalysis. Enzyme and Microbial Technology 41, 628–637.

Kazlauskas, R.J., Bornscheuer, U.T., 2009. Finding better protein engineering strategies. Nature chemical biology 5, 526–9.

Kemp, D.S., O'Brien, J.T., 1970. Base catalysis of thiazolium salt hydrogen exchange and its implications for enzymic thiamine cofactor catalysis. Journal of the American Chemical Society 92, 2554–2555.

Kern, D., Kern, G., Neef, H., Tittmann, K., Killenberg-Jabs, M., Wikner, C., Schneider, G., Hübner, G., 1997. How thiamine diphosphate is activated in enzymes. Science 275, 67–70.

Kluger, R., Tittmann, K., 2008. Thiamin diphosphate catalysis: enzymic and nonenzymic covalent intermediates. Chemical reviews 108, 1797–833.

Kobori, Y., Myles, D.C., Whitesides, G.M., 1992. Substrate specificity and carbohydrate synthesis using transketolase. The Journal of Organic Chemistry 57, 5899–5907.

Kochetov, G.A., Sevostyanova, I.A., 2005. Binding of the coenzyme and formation of the transketolase active center. IUBMB Life 57, 491–7.

Koeller, K.M., Wong, C.H., 2001. Enzymes for chemical synthesis. Nature 409, 232–40.

Koszelewski, D., Tauber, K., Faber, K., Kroutil, W., 2010. omega-Transaminases for the synthesis of non-racemic alpha-chiral primary amines. Trends in biotechnology28, 324–32.

Kuchner, O., Arnold, F.H., 1997. Directed evolution of enzyme catalysts. Trends in biotechnology 15, 523–30.

Langenbeck, W., 1932. Fermentproblem und organische Katalyse. Angewandte Chemie 45, 97–99.

Lee, J., Goodey, N.M., 2011. Catalytic contributions from remote regions of enzyme structure.Chemical reviews 111, 7595–624.

Leemhuis, H., Kelly, R.M., Dijkhuizen, L., 2009. Directed evolution of enzymes: Library screening strategies. IUBMB Life 61, 222–8.

Lehmann, M., Wyss, M., 2001. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. Current opinion in biotechnology 12, 371–5.

Lindqvist, Y., Schneider, G., 1993. Thiamin diphosphate dependent enzymes: transketolase, pyruvate oxidase and pyruvate decarboxylase. Current Opinion in Structural Biology 3, 896–901.

Lockless, S.W., Ranganathan, R., 1999. Evolutionarily conserved pathways of energetic connectivity in protein families. Science 286, 295–9.

Lutz, S., 2010. Beyond directed evolution--semi-rational protein engineering and design. Current opinion in biotechnology 21, 734–43.

Lutz, S., Ostermeier, M., Moore, G.L., Maranas, C.D., Benkovic, S.J., 2001. Creating multiple-crossover DNA libraries independent of sequence identity. Proceedings of the National Academy of Sciences of the United States of America 98, 11248–53.

Machajewski, T., Wong, C., 2000. The Catalytic Asymmetric Aldol Reaction. Angewandte Chemie (International ed. in English) 39, 1352–1375.

Martinez-Torres, R.J., Aucamp, J.P., George, R., Dalby, P.A., 2007. Structural stability of E. coli transketolase to urea denaturation. Enzyme and Microbial Technology 41, 653–662.

Meshalkina, L., Nilsson, U., Wikner, C., Kostikowa, T., Schneider, G., 1997. Examination of the thiamin diphosphate binding site in yeast transketolase by site-directed mutagenesis. European journal of biochemistry / FEBS 244, 646–52.

Mildvan, A.S., Weber, D.J., Kuliopulos, A., 1992. Quantitative interpretations of double mutations of enzymes. Archives of biochemistry and biophysics 294, 327–40.

Miller, O.J., Hibbert, E.G., Ingram, C.U., Lye, G.J., Dalby, P.A., 2007. Optimisation and evaluation of a generic microplate-based HPLC screen for transketolase activity. Biotechnology letters 29, 1759–70.

Mizuhara, S., Handler, P., 1954. Mechanism of Thiamine-catalyzed Reactions. Journal of the American Chemical Society 76, 571–573.

Morley, K.L., Kazlauskas, R.J., 2005. Improving enzyme properties: when are closer mutations better? Trends in biotechnology 23, 231–7.

Morris, K.G., Smith, M.E.B., Turner, N.J., Lilly, M.D., Mitra, R.K., Woodley, J.M., 1996. Transketolase from Escherichia coli: A practical procedure for using the biocatalyst for asymmetric carbon-carbon bond synthesis. Tetrahedron: Asymmetry 7, 2185–2188.

Mutti, F.G., Fuchs, C.S., Pressnitz, D., Sattler, J.H., Kroutil, W., 2011. Stereoselectivity of Four (R)-Selective Transaminases for the Asymmetric Amination of Ketones.Advanced Synthesis & Catalysis 353, 3227–3233.

Myles, D.C., Andrulis, P.J., Whitesides, G.M., 1991. A transketolase-based synthesis of (+)-exo-brevicomin. Tetrahedron Letters 32, 4835–4838.

Nagabhushan, T., Miller, G.H., Varma, K.J., 2000. Chloramphenicol and Analogues, in: Kirk-Othmer Encyclopedia of Chemical Technology. John Wiley & Sons, Inc.

Neher, E., 1994. How frequent are correlated changes in families of protein sequences? Proceedings of the National Academy of Sciences of the United States of America 91, 98–102.

Nemeria, N.S., Chakraborty, S., Balakrishnan, A., Jordan, F., 2009. Reaction mechanisms of thiamin diphosphate enzymes: defining states of ionization and tautomerization of the cofactor at individual steps. The FEBS journal 276, 2432–2446.

Nikkola, M., Lindqvist, Y., Schneider, G., 1994. Refined structure of transketolase from Saccharomyces cerevisiae at 2.0 A resolution. Journal of molecular biology 3, 387-404.

Nilsson, U., Hecquet, L., Gefflaut, T., Guerard, C., Schneider, G., 1998. Asp477 is a determinant of the enantioselectivity in yeast transketolase. FEBS letters 424, 49–52.

Nilsson, U., Meshalkina, L., Lindqvist, Y., Schneider, G., 1997. Examination of substrate binding in thiamin diphosphate-dependent transketolase by protein crystallography and site-directed mutagenesis. The Journal of biological chemistry 272, 1864–1869.

Ostermeier, M., Shim, J.H., Benkovic, S.J., 1999. A combinatorial approach to hybrid enzymes independent of DNA homology. Nature biotechnology 17, 1205–9.

Otten, L.G., Hollmann, F., Arends, I.W.C.E., 2010. Enzyme engineering for enantioselectivity: from trial-and-error to rational design? Trends in biotechnology 28, 46–54.

Otten, L.G., Quax, W.J., 2005. Directed evolution: selecting today's biocatalysts. Biomolecular engineering 22, 1–9.

Paramesvaran, J., Hibbert, E.G., Russell, A.J., Dalby, P.A., 2009. Distributions of enzyme residues yielding mutants with improved substrate specificities from two different directed evolution strategies. Protein engineering, design & selection : PEDS 22, 401–11.

Park, E.-S., Kim, M., Shin, J.-S., 2012. Molecular determinants for substrate selectivity of ω-transaminases. Applied microbiology and biotechnology 6, 2425-35.

Patel, R.N., 2008. Synthesis of chiral pharmaceutical intermediates by biocatalysis. Coordination Chemistry Reviews 252, 659–701.

Patrick, W.M., Firth, A.E., Blackburn, J.M., 2003. User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. Protein engineering 6, 451-457

Pavelka, A., Chovancova, E., Damborsky, J., 2009. HotSpot Wizard: a web server for identification of hot spots in protein engineering. Nucleic acids research 37, W376–83.

Payongsri, P., Steadman, D., Strafford, J., MacMurray, A., Hailes, H.C., Dalby, P.A., 2012. Rational substrate and enzyme engineering of transketolase for aromatics. Organic & biomolecular chemistry 10, 9021–9.

Poole, A.M., Ranganathan, R., 2006. Knowledge-based potentials in protein design. Current opinion in structural biology 16, 508–13.

Reetz, M.T., Bocola, M., Carballeira, J.D., Zha, D., Vogel, A., 2005. Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. Angewandte Chemie (International ed. in English) 44, 4192–6.

Reetz, M.T., Carballeira, J.D., 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. Nature protocols 2, 891–903.

Resch, V., Schrittwieser, J.H., Siirola, E., Kroutil, W., 2011. Novel carbon-carbon bond formations for biocatalysis. Current opinion in biotechnology 22, 793–799.

Rios-Solis, L., Halim, M., Cázares, A., Morris, P., Ward, J.M., Hailes, H.C., Dalby, P.A., Baganz, F., Lye, G.J., 2011. A toolbox approach for the rapid evaluation of multi-step enzymatic syntheses comprising a “mix and match” E. coli expression system with microscale experimentation. Biocatalysis and Biotransformation 29, 192–203.

Robertson, D.E., Steer, B.A., 2004. Recent progress in biocatalyst discovery and optimization. Current opinion in chemical biology 8, 141–9.

Romero, P.A., Arnold, F.H., 2009. Exploring protein fitness landscapes by directed evolution. Nature reviews. Molecular cell biology 10, 866–76.

Röthlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E. a, Zanghellini, A., Dym, O., Albeck, S., Houk, K.N., Tawfik, D.S., Baker, D., 2008. Kemp elimination catalysts by computational enzyme design. Nature 453, 190–5.

Rudat, J., Brucher, B.R., Syldatk, C., 2012. Transaminases for the synthesis of enantiopure beta-amino acids. AMB Express 2, 11.

Samland, A.K., Sprenger, G. a, 2006. Microbial aldolases as C-C bonding enzymes--unknown treasures and new developments. Applied microbiology and biotechnology 71, 253–64.

Savile, C.K., Janey, J.M., Mundorff, E.C., Moore, J.C., Tam, S., Jarvis, W.R., Colbeck, J.C., Krebber, A., Fleitz, F.J., Brands, J., Devine, P.N., Huisman, G.W., Hughes, G.J., 2010. Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. Science329, 305-309.

Schell, U., Wohlgemuth, R., Ward, J.M., 2009. Synthesis of pyridoxamine 5’-phosphate using an MBA:pyruvate transaminase as biocatalyst. Journal of Molecular Catalysis B: Enzymatic 59, 279–285.

Schellenberger, a, 1998. Sixty years of thiamin diphosphate biochemistry. Biochimica et biophysica acta1385, 177–86.

Schenk, G., Duggleby, R.G., Nixon, P.F., 1998. Properties and functions of the thiamin diphosphate dependent enzyme transketolase. The international journal of biochemistry & cell biology 30, 1297–318.

Schenk, G., Layfield, R., Candy, J.M., Duggleby, R.G., Nixon, P.F., 1997. Molecular evolutionary analysis of the thiamine-diphosphate-dependent enzyme, transketolase. Journal of molecular evolution 44, 552–72.

Schneider, G., Fiedler, E., Thorell, S., Sandalova, T., Golbik, R., Ko, S., 2002. Snapshot of a key intermediate in enzymatic thiamin catalysis : Crystal structure of the

alpha,betacarbanion of the active site of transketolase from Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America 99, 591–595.

Schneider, G., Lindqvist, Y., 1998. Crystallography and mutagenesis of transketolase: mechanistic implications for enzymatic thiamin catalysis. Biochimica et biophysica acta 1385, 387–98.

Schreiber, G., Fersht, A.R., 1995. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. Journal of molecular biology 248, 478–86.

Shin, J.S., Kim, B.G., 2001. Comparison of the omega-transaminases from different microorganisms and application to production of chiral amines. Bioscience, biotechnology, and biochemistry 65, 1782-1788.

Shin, J.-S., Kim, B.-G., 2002. Exploring the active site of amine:pyruvate aminotransferase on the basis of the substrate structure-reactivity relationship: how the enzyme controls substrate specificity and stereoselectivity. The Journal of organic chemistry 9, 2848-2853.

Shin, J.-S., Yun, H., Jang, J.-W., Park, I., Kim, B.-G., 2003. Purification, characterization, and molecular cloning of a novel amine:pyruvate transaminase from Vibrio fluvialis JS17. Applied microbiology and biotechnology61, 463–71.

Sieber, V., Martinez, C.A., Arnold, F.H., 2001. Libraries of hybrid proteins from distantly related sequences. Nature biotechnology 19, 456–60.

Singleton, C.K., Wang, J.J., Shan, L., Martin, P.R., 1996. Conserved residues are functionally distinct within transketolases of different species. Biochemistry 35, 15865–9.

Smith, J.M., 1970. Natural selection and the concept of a protein space. Nature 225, 563–4.

Smith, M.E.B., Chen, B.H., Hibbert, E.G., Kaulmann, U., Smithies, K., Galman, J.L., Baganz, F., Dalby, P.A., Hailes, H.C., Lye, G.J., Ward, J.M., Woodley, J.M., Micheletti, M., 2010. A Multidisciplinary Approach Toward the Rapid and Preparative-Scale Biocatalytic Synthesis of Chiral Amino Alcohols: A Concise Transketolase-/ω-Transaminase-Mediated Synthesis of (2 S ,3 S)-2-Aminopentane-1,3-diol. Organic Process Research & Development 14, 99–107.

Smith, M.E.B., Hibbert, E.G., Jones, A.B., Dalby, P.A., Hailes, H.C., 2008. Enhancing and Reversing the Stereoselectivity of Escherichia coli Transketolase via Single-Point Mutations. Organic Process Research & Development 350, 2631–2638.

Smith, M.E.B., Smithies, K., Senussi, T., Dalby, P.A., Hailes, H.C., 2006. The First Mimetic of the Transketolase Reaction. European Journal of Organic Chemistry 2006, 1121–1123.

Smithies, K., Smith, M.E.B., Kaulmann, U., Galman, J.L., Ward, J.M., Hailes, H.C., 2009. Stereoselectivity of an ω-transaminase-mediated amination of 1,3-dihydroxy-1-phenylpropane-2-one. Tetrahedron: Asymmetry 20, 570–574.

Socolich, M., Lockless, S.W., Russ, W.P., Lee, H., Gardner, K.H., Ranganathan, R., 2005. Evolutionary information for specifying a protein fold. Nature 437, 512–8.

Sprenger, G.A., Pohl, M., 1999. Synthetic potential of thiamin diphosphate-dependent enzymes. Journal of Molecular Catalysis B: Enzymatic 6, 145–159.

Sprenger, G.A., Schörken, U., Sprenger, G., Sahm, H., 1995. Transketolase A of Escherichia coli K12. Purification and properties of the enzyme from recombinant strains. European journal of biochemistry / FEBS 230, 525–32.

Stemmer, W.P., 1994a. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. Proceedings of the National Academy of Sciences of the United States of America 91, 10747–51.

Stemmer, W.P., 1994b. Rapid evolution of a protein in vitro by DNA shuffling. Nature370, 389–91.

Strafford, J., Payongsri, P., Hibbert, E.G., Morris, P., Batth, S.S., Steadman, D., Smith, M.E.B., Ward, J.M., Hailes, H.C., Dalby, P.A., 2012. Directed evolution to re-adapt a co-evolved network within an enzyme.Journal of biotechnology 157, 237–245.

Street, A.G., Mayo, S.L., 1999. Computational protein design. Structure (London, England : 1993) 7, R105–9.

Süel, G.M., Lockless, S.W., Wall, M. a, Ranganathan, R., 2003. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nature structural biology 10, 59–69.

Sundström, M., Lindqvist, Y., Schneider, G., 1992. Three-dimensional structure of apotransketolase. Flexible loops at the active site enable cofactor binding. FEBS letters 313, 229–31.

Takayama, S., McGarvey, G.J., Wong, C.-H., 1997. Enzymes in organic synthesis: recent developments in aldol reactions and glycosylations. Chemical Society Reviews 26, 407.

Thomas, J.G., Baneyx, F., 1996. Protein misfolding and inclusion body formation in recombinant Escherichia coli cells overexpressing Heat-shock proteins. The Journal of biological chemistry271, 11141–7.

Thomas, J.G., Baneyx, F., 1997. Divergent effects of chaperone overexpression and ethanol supplementation on inclusion body formation in recombinant Escherichia coli. Protein expression and purification 11, 289–96.

Tokuriki, N., Stricher, F., Serrano, L., Tawfik, D.S., 2008. How protein stability and new functions trade off. PLoS computational biology 4, e1000002.

Tokuriki, N., Tawfik, D.S., 2009. Stability effects of mutations and protein evolvability. Current opinion in structural biology 19, 596–604.

Toone, E.J., Whitesides, G.M., 1991. Enzymes as Catalysts in Carbohydrate Synthesis., in: Symposium A Quarterly Journal In Modern Foreign Literatures. pp. 1–22.

Toscano, M.D., Woycechowsky, K.J., Hilvert, D., 2007. Minimalist active-site redesign: teaching old enzymes new tricks. Angewandte Chemie (International ed. in English) 46, 3212–36.

Tufvesson, P., Lima-Ramos, J., Jensen, J.S., Al-Haque, N., Neto, W., Woodley, J.M., 2011. Process considerations for the asymmetric synthesis of chiral amines using transaminases. Biotechnology and bioengineering 108, 1479–93.

Turner, N.J., 2000. Applications of transketolases in organic synthesis.Current opinion in biotechnology 11, 527–31.

Turner, N.J., 2009. Directed evolution drives the next generation of biocatalysts. Nature chemical biology 5, 567–73.

Wang, W., Nema, S., Teagarden, D., 2010. Protein aggregation--pathways and influencing factors. International journal of pharmaceutics390, 89–99.

Ward, J., Wohlgemuth, R., 2010. High-Yield Biocatalytic Amination Reactions in Organic Synthesis. Current Organic Chemistry 14, 1914–1927.

Wells, J.A., 1990. Additivity of mutational effects in proteins. Biochemistry 29, 8509–8517.

Wetzel, R., 1994. Mutations and off-pathway aggregation of proteins. Trends in biotechnology 12, 193–8.

Wiesner, K., Valenta, Z., 1956. The mechanism of reactions catalyzed by thiamine pyrophosphate. Experientia 12, 190–193.

Wikner, C., Meshalkina, L., Nilsson, U., Bäckström, S., Lindqvist, Y., Schneider, G., 1995. His103 in yeast transketolase is required for substrate recognition and catalysis. European journal of biochemistry / FEBS 233, 750–5.

Wikner, C., Nilsson, U., Meshalkina, L., Udekwu, C., Lindqvist, Y., Schneider, G., 1997. Identification of catalytically important residues in yeast transketolase. Biochemistry 36, 15643–9.

Wohlgemuth, R., 2009. C2-Ketol elongation by transketolase-catalyzed asymmetric synthesis. Journal of Molecular Catalysis B: Enzymatic 61, 23–29.

Wu, S., Acevedo, J.P., Reetz, M.T., 2010. Induced allostery in the directed evolution of an enantioselective Baeyer-Villiger monooxygenase. Proceedings of the National Academy of Sciences of the United States of America 107, 2775–80.

Yang, G., Withers, S.G., 2009. Ultrahigh-throughput FACS-based screening for directed enzyme evolution. Chembiochem : a European journal of chemical biology 10, 2704–15.

Yi, D., Devamani, T., Abdoul-Zabar, J., Charmantray, F., Helaine, V., Hecquet, L., Fessner, W.-D., 2012. A pH-based high-throughput assay for transketolase: fingerprinting of

substrate tolerance and quantitative kinetics. Chembiochem : a European journal of chemical biology 13, 2290–300.

Yonaha, K., Toyama, S., Kagamiyama, H., 1983. Properties of the bound coenzyme and subunit structure of omega-amino acid:pyruvate aminotransferase. The Journal of biological chemistry 258, 2260–5.

Yonaha, K., Toyama, S., Yasuda, M., Soda, K., 1976. Purification and crystallization of bacterial ω-amino acid-pyruvate aminotransferase. FEBS Letters 71, 21–24.

You, L., Arnold, F.H., 1996. Directed evolution of subtilisin E in Bacillus subtilis to enhance total activity in aqueous dimethylformamide. Protein engineering 9, 77–83.

Zhao, H., Giver, L., Shao, Z., Affholter, J.A., Arnold, F.H., 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. Nature biotechnology 16, 258–61.

Ziegler, T., Straub, A., Effenberger, F., 1988. Enzyme-Catalyzed Synthesis of 1-Deoxymannojirimycin, 1-Deoxynojirimycin, and 1,4-Dideoxy-1,4-imino-D-arabinitol. Angewandte Chemie International Edition in English 27, 716–717.

9.3. Second Generation Engineering of Transketolase for Polar Aromatic Aldehyde Substrates

Second Generation Engineering of Transketolase for Polar Aromatic Aldehyde Substrates

Panwajee Payongsri[a], David Steadman[b], Helen C. Hailes[b], Paul A. Dalby[a,*]

[a]Advanced Centre for Biochemical Engineering, Department of Biochemical Engineering, University College London, Torrington Place, London, WC1E 7JE, UK.

[b]Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, UK.

[*]Corresponding author

email address: p.dalby@ucl.ac.uk

Abbreviation

3-FBA, 3-formylbenzoic acid; 3-HBA, 3-hydroxybenzaldehyde; 4-FBA, 4-formylbenzoic acid; HPA, hydroxypyruvic acid; TFA, trifluoroacetic acid; ThDP, thiamine diphosphate; TK, transketolase

Abstract

Transketolase mutants evolved for propanal activity were previously found to have nascent activity towards polar aromatic aldehydes 3-formylbenzoic acid (3-FBA), 4-formylbenzoic acid (4-FBA), and 3-hydroxybenzaldehyde (3-HBA). We have now evolved transketolase further for improved aromatic aldehyde activities, using saturation mutagenesis at two active-site residues predicted to interact with the aromatic substituent moieties. Mutations at S385 gave up to 13-fold enhanced activities, whereas mutations at R358 did not improve the activity towards any substrate. One particular S385 mutant completely removed the substrate inhibition by 3-FBA, observed in all previous mutants. Mutations at S385 selectively altered the preference of transketolase towards each of the three substrates. Interestingly, the mechanisms of catalytic improvement were dependent both on mutation type and substrate. S385E improved 3-FBA activity mostly via k_{cat}, but reduced 4-FBA activity mostly via K_M. Conversely, S385Y/T improved 3-FBA activity mostly via K_M, and improved 4-FBA activity mostly via k_{cat}. This suggested that substrate proximity and orientation in the active site is very sensitive to mutation. Comparison of all mutant activities obtained for each substrate, and substrates computationally docked into mutant active-sites, indicated divergent binding-modes for the three aromatic substrates. Hence further improvements towards each substrate may now require different saturation mutagenesis strategies.

Keywords

Biocatalysis, transketolase, enzyme engineering, benzaldehyde, directed evolution

1. Introduction

Asymmetric carbon-carbon bond formation is a powerful method in organic synthesis (Breuer and Hauer, 2003; Enders and Narine, 2008; Fessner, 1998; Resch et al., 2011; Wohlgemuth, 2009). The ability to perform new asymmetric carbon-carbon formation can open up routes to a wide range of novel and natural compounds which could serve as a building block for further synthesis (Enders and Narine, 2008). Enzymes have gained particular popularity in this area due to their high selectivity and specificity. Several of these enzymes, including transketolase (TK), have been shown to be able to catalyse asymmetric carbon-carbon bond formation with considerable synthetic potential (Breuer and Hauer, 2003; Enders and Narine, 2008; Fessner, 1998; Fessner and Helaine, 2001; Morris et al., 1996; Resch et al., 2011; Samland and Sprenger, 2006; Takayama et al., 1997; Toone and Whitesides, 1991).

Transketolase (EC 2.2.1.1), belongs to the thiamine diphosphate (ThDP) dependent enzyme family, and plays two crucial roles in the non-oxidative pentose phosphate pathway and Calvin cycle (Sprenger et al., 1995). It catalyses a reversible transfer of a C_2-hydroxyketone group from a ketol donor such as ketose sugars, to an aldehyde (Schenk et al., 1998; Schneider and Lindqvist, 1998; Schörken and Sprenger, 1998). Transketolase catalysed reactions produce an α,α-dihydroxy ketone group which is found in a wide range of natural compounds such as carbohydrates and corticosteroids (Hailes et al., 2010). Hydroxypyruvate has been extensively used as a ketol donor instead of the natural ketose sugar substrates, due to the production of carbon dioxide, which can drive the reaction to completion. The use of transketolase could provide advantages over chemical synthesis routes which tend to suffer from low yields due to multistep procedures (Hailes et al., 2010) and racemic products or low stereoselectivities when a single step method is applied (Galman et al., 2012; Smith et al., 2006). So far, transketolases from several organisms have been exploited in new synthetic routes to carbohydrates and sugar analogues (Bolte et al., 1987; Dalmas and Demuynck, 1993; Demuynck et al., 1991; Kobori et al., 1992), (+)-*exo*-brevicomin (Myles et al., 1991), N- hydroxypyrrolidine (Humphrey et al., 2000), furaneol (Hecquet et al., 1996), the glycosidase inhibitor 1,4-dideoxy-1,4-imino-D-arabinitol (Ziegler et al., 1988), and intermediates for the synthesis of chiral amino alcohols (Ingram et al., 2007).

Although wild-type transketolases can accept a wide range of aliphatic, heteroaromatic, and cyclic aldehydes (Bolte et al., 1987; Chen et al., 2008; Demuynck et al., 1991; Morris et al., 1996; Schörken and Sprenger, 1998), short chain aliphatic aldehydes tend to give rise to faster reactions than for more sterically challenging substrates (Bolte et al., 1987; Chen et al., 2008; Demuynck et al., 1991). In addition, most of the synthetic products previously reported have used aliphatic hydroxylated aldehydes as starting materials. Expansion of the substrate range accepted by transketolases will increase their potential use for the synthesis of novel dihydroxy ketone compounds. Indeed, protein engineering of WT *E. coli* TK has been successfully used to improve its activity towards aliphatic aldehydes (Hibbert et al., 2008), with both enhanced and reversed stereospecificity (Cazares et al., 2010; Galman et al., 2010; Smith et al., 2008; Yi et al., 2012). By contrast, aromatic aldehydes still suffer from low reaction yields (Galman et al., 2010), except 3-formylbenzoic acid (3-FBA) and 4-formylbenzoic acid (4-FBA) which introduce a beneficial binding interaction between their carboxylic acid group and the transketolase phosphate binding residues (Payongsri et al., 2012). The *E. coli* TK variant D469T was found to have the highest activity towards the two carboxylated aromatic substrates 3-FBA and 4-FBA, from all the previously constructed mutants tested (Payongsri et al., 2012). However, in all cases the activity was still low compared to any natural substrates, and for 3-FBA, substrate inhibition was also always present.

Saturation mutagenesis at residues that interact directly with substrates have the greatest potential to improve enzyme activity and stereospecificity in a single round of evolution, compared to mutation of residues at more distant locations (Paramesvaran et al., 2009). Here, we have now applied a targeted saturation mutagenesis strategy to further improve the activity and yield towards the aromatic aldehyde substrates 3-FBA, 4-FBA and additionally, 3-hydroxybenzaldehyde (3-HBA).

Previously, the mutant D469T/R520Q was found to retain most of the activity of D469T, towards 3-FBA and 4-FBA. The saturation mutagenesis was targeted into D469T/R520Q, as it is also more readily able to accommodate further mutations due to stabilisation of the enzyme in terms of protection from the formation of insoluble aggregates (Strafford et al., 2012). Residue S385 is within the second cofactor binding loop of *E. coli* transketolase (Martinez-Torres et al., 2007). The side chain hydroxyl group of S385 can form a hydrogen bond with the phosphate moiety of natural substrates, as determined from several *E. coli* TK crystal structures with substrates (2R5N, 2R8O, 2R8P) (Asztalos et al., 2007). Residue R358 is also involved in binding the phosphate moiety of natural substrates (Asztalos et al., 2007), as confirmed by selected mutagenesis of the yeast TK (Nilsson et al., 1997). Previous mutations, R358P and R358L also indicated an analogous interaction with the carboxylate moieties of 3-FBA and 4-FBA (Payongsri et al., 2012). In our previous work, saturation mutagenesis of S385 and R358, along with 18 other sites within the wild-type enzyme, aimed to improve activity towards non-aromatic aldehydes, glycolaldehyde and propanal (Hibbert et al., 2007; Hibbert et al., 2008). While several other sites gave significantly improved activities, S385 gave no mutants of interest, and R358 yielded only R358I with a 2-fold improvement of activity towards each substrate. Here, these sites have been mutated to all possible amino acids, but within D469T/R520Q, with the aim of further modifying the activity towards substituted aromatic aldehydes.

2. Materials and Methods

2.1. Chemicals and reagents

All chemical reagents were purchased from Sigma-Aldrich (Aldrich® Chemistry, UK), otherwise stated. Lithium hydroxypyruvate was prepared according to the previous protocol (Morris et al., 1996).

2.2. Transketolase library

All mutations were constructed on D469T/R520Q due to its stability (Strafford et al., 2012), and initial activity on the aromatic aldehydes (Payongsri et al., 2012). The transketolase D469T/R520Q mutant gene and all further mutants were expressed under the control of *tktA* gene promoter in the plasmid pQR791 in XL-10 Gold (Stratagene) (Martinez-Torres et al., 2007). Target residues for saturation mutagenesis were from those determined to be in the first shell (Hibbert et al., 2007). Mutagenesis was targeted only to those involved in binding phosphate groups within natural substrates, as determined by previous mutagenesis and crystal structure analysis (Asztalos et al., 2007), as equivalent interactions with substituted aromatic aldehydes were previously found by us to greatly influence activity (Payongsri et al., 2012; Strafford et al., 2012). These included S385, R358, H461 and R520. R520 was already mutated to R520Q in the parent variant D469T/R520Q, and so not mutated further. H461 was also excluded from further mutation as it was previously found to lead to stability issues (Payongsri et al., 2012; Strafford et al., 2012). Mutagenesis was carried out using Quikchange site directed mutagenesis (Stratagene). The *dpnI*-digested PCR product was transformed into XL-10 gold competent cells (Stratagene). The quality and diversity of the PCR was checked by DNA sequencing prior screening. The primers used for additional mutant construction were listed below (codon underlined, mutations in bold, and N is an equal mix of all four bases, H is 50% A and 50% T, and S is 50% G and 50% C).

R358X GAAAATCGCCAGC**NHN**AAAGCGTCTCAGAATG

S385X GCTGACCTGGCGCCG**NHS**AACCTGACCCTGTGG

2.3. Library Screening

After the transformation, 24 colonies for each library were inoculated into 12 ml LB media with 150 µg/ml ampicillin, incubated for 18 hours at 37 °C, shaking at 250 rpm and then 2 ml of the culture was centrifuged at 13000 rpm for 10 minutes. The probability of coverage and possible number of amino acid replacements was calculated using the online tool: GLUE1AA (Patrick et al., 2003). From each codon mix used, the two libraries could produce 16 possible variants and so 11.6

distinct amino acids were expected to be obtained by picking 24 colonies. The diversity of mutations was confirmed by DNA sequencing of 6 random colonies from each library. The cell pellets were resuspended in 200 µl of 2x cofactor solution (4.8 mM ThDP and 18 mM $MgCl_2$ in 50 mM Tris.HCl, pH 7.0), incubated for 20 minutes and 150 µl then transferred to a borosilicate microplate (Radleys, Essex, UK). Reactions were started by the addition of 150 µl of 2x substrate solution (100 mM 3-FBA and 100 mM HPA, 80 mM 4-FBA and 80 mM HPA, or 30mM 3-HBA and 60mM HPA, each in 50 mM Tris buffer, pH 7.0). All reactions were performed at 22 °C with shaking at 200 rpm, 3 mm amplitude, to avoid cell sedimentation, and sealed to prevent evaporation (Thermo Scientific Nunc). 3-FBA samples were taken at 1 hour and 24 hours after the reaction was started to screen for both activity (conversion at 1 hour) and final yield at 24 hours, based on the product concentration. Previous mutants had given low activities and yields towards 4-FBA and 3-HBA, where the yields were reduced in proportion to their activities due to competing degradation of HPA, catalysed by D469T containing variants (Payongsri et al., 2012). Therefore, the libraries were screened for conversion after 18 hours reaction to identify mutants with improved final yields for 4-FBA and 3-HBA. At each time-point, 20 µl of the reaction sample was added to 380 µl 0.1% TFA, centrifuged at 13000 rpm for 3 mins, and the supernatants analysed by HPLC with an ACE5 C18 reverse phase column (150x4.6 mm). The 3-FBA and 4-FBA reaction samples were analysed as previously described (Payongsri et al., 2012) using two mobile phases 0.1% TFA and 100% acetonitrile, and a flow rate of 1 ml/min. The flow was started at 85% TFA (0.1%) and reduced to 28% over 9 mins, followed by 2 mins equilibration at 85% TFA (0.1%) and 15% acetonitrile. The retention times of 3-(1,3-Dihydroxy-2-oxopropyl) benzoic acid, 3-FBA, 4-(1,3-Dihydroxy-2-oxopropyl) benzoic acid and 4-FBA were 2.97, 5.47, 2.45 and 5.41 mins, respectively. The 3-HBA reaction samples were analysed by gradient elution as follows. Using a mobile phase of 0.2M acetic acid and 80% methanol at the flow rate of 1ml/min, the flow profile was separated into 3 phases: 1st phase 5 mins using 90% 0.2 M acetic acid/10% methanol (80%v/v); 2nd phase ramped linearly to 40% 0.2 M acetic acid/60% methanol (80%v/v) over 14 mins, and the last phase maintaining 40% 0.2 M acetic acid/60% methanol (80%v/v) for 6 mins. The column was then re-equilibrated for 3 mins using 90% 0.2 M acetic acid/10% methanol (80%v/v). The retention time of the product from the 3-HBA reaction was 4.7 min and that for 3-HBA was 14.5 min. Two colonies with each of low, medium and enhanced activities towards each substrate were picked and their activities assessed in more detail using clarified lysate to validate the screening results.

2.4. Cell culture and protein quantification for detailed enzyme kinetics

Glycerol stocks of selected mutants were re-streaked on 150 µg/ml ampicillin-LB plates and incubated at 37 °C for 18 hours. Single colonies were inoculated into 20 ml LB with 150 µg/ml ampicillin in 250 ml shake flasks, incubated at 37 °C with shaking at 250 rpm for 18 hours, then harvested by centrifugation. Supernatant-free cell pellets were resuspended in Tris buffer, pH 7.0 and sonicated on ice (MSE Soniprep 150 probe, Sanyo) with 10s on, 15s off for 10 cycles. Cell debris was removed by centrifugation at 17,700 g for 10 minutes at 4 °C. The clarified lysate was aliquotted and stored at -80 °C. Total protein concentration in the clarified lysate was determined using the Bradford assay with BSA as a standard protein. The lysate was further analysed by SDS-PAGE and densitometry as previously described (Hibbert et al., 2007) to determine the TK concentration which is always over-expressed to above 20% of the total protein. The plasmids of these mutants were extracted and sequenced to identify the mutation.

2.5. Detailed enzyme kinetics

Clarified lysates of the mutants with high, moderate, and low enzyme activities or conversion yields, prepared as above, were used to determine their specific activities or yield towards each substrate. Mutants with high activities towards at least one substrate were further studied to determine the K_M and k_{cat} for all the substrates. Kinetic parameters were obtained at saturating 50 mM HPA levels and the aldehyde concentrations varied (3-FBA 3 to 90 mM, 4-FBA 6 to 30 mM, 3-HBA 5 to 25 mM). Substrate solutions were prepared at 3x in 50 mM Tris.HCl, pH 7.0. To 30 µl of clarified lysate, 30 µl of 10x cofactor solution (24 mM ThDP, 90 mM $MgCl_2$) was added with 140 µl 50 mM Tris.HCl, pH 7.0, and incubated for 20 mins. Reactions were initiated by the addition of 100 µl of substrate solution. All reactions were carried out in triplicate in glass vials at 22 °C. To 20 µl of reaction samples, 380 µl of 0.1% TFA was added and the mixture centrifuged at 13,000 rpm for 3 mins. The supernatants were analysed by HPLC as above. 3-FBA reaction samples were taken at every 3 mins for 15 mins. The reaction samples of 4-FBA were taken at 15 minute intervals for 90 mins. 3-HBA reaction samples were taken every 30 mins for 180 mins. Samples at 18 hours for the highest aldehyde concentrations were also taken to quantify the reaction yield determined from product concentrations. The TK concentration in each reaction was between 0.07 and 0.3 mg/ml. Higher TK concentrations were used for the 3-HBA reaction due to the slower conversion. All data were fitted by non-linear regression to the Michaelis-Menten equation to determine the K_M and k_{cat} of all the mutants with each substrate. Double-reciprocal Lineweaver-Burk plots were also used to verify that the relationships between the velocity and the concentration were linear. The racemic

products from 3-HBA (Galman et al., 2010), 3-FBA and 4-FBA (Payongsri et al., 2012) were synthesised, purified, and characterised as previously, as standards for HPLC calibration.

2.6. Computational modelling of 3-FBA, 4-FBA and 3-HBA binding into transketolase mutant active site

The structures of TK mutants, aldehydes and ThDP enamine were modelled as previously described (Payongsri et al., 2012). TK mutant structures were prepared by replacing the target residues of the *E. coli* wild type (PDB ID: 1QGD). Residues within 10 Å from the ThDP molecule were selected and subjected to energy minimised by Discovery Studio 2.0 (Accelrys, Inc. San Diego, California, USA). The energy minimised was performed for 1000 steps by Charm Forcefield, Adopted Basic NR, Implicit Generalised Born solvent model, True SHAKE constant (Payongsri et al., 2012).

ThDP-enamine and all the aldehyde substrates were drawn in chem3D ultra v.10. They were then energy minimised by MM2 calculation. The energy minimised ThDP-enamine structure was then docked into the active site of the energy-minimised TK mutants prepared as above using. The aldehydes were then docked into these ThDP-enamine-TK structures. All computational docking were performed by AutoDock 4.2 and AutoDock Tools 1.5.4 (Goodsell et al., 1996). The centre of all the mutants for docking the aldehydes was 11.777, 27.078, 37.195 with the grid size of 30 Å × 30 Å × 30 Å.

3. Results and Discussion

3.1. High-throughput screening validation using whole cells and cell lysates

The screening was performed in whole cells to avoid uneven cell disruption which can influence the amount of TK released and the bioconversion rate and yield. To confirm that the whole cell can represent the clarified lysate, their yields and activities were compared. For 4-FBA and 3-HBA, the conversion yields based on product formed after 18 hours reaction with whole cells, were screened at high throughput as an indirect measure of activity, as it maximised the signal to noise. Our previous best mutants, D469T and D469T/R520Q, gave very low activities and yields towards 4-FBA and 3-HBA, and for previous D469T-containing mutants all final yields obtained were found to be proportional to their specific activities (Payongsri et al., 2012). This was due to competition with an irreversible HPA degradation reaction catalysed by the D469T-containing variants, and was not due to any background from host-cell enzymes or the buffer (Payongsri et al., 2012).

After screening all three aromatic aldehyde substrates on both the S385X/D469T/R520Q and R358X/D469T/R520Q libraries with whole cells, mutants were selected randomly across the range from low to high activity for a more detailed analysis of specific activities and yields obtained with clarified lysates instead of whole cells. The correlation between specific activity and final yield after 18 hours remained true for 4-FBA and 3-HBA with five mutants selected randomly from the new library (see Supplementary Information Figure S1).

For 3-FBA, reactions were considerably faster than for 4-FBA or 3-HBA, and so the background HPA degradation was much less relevant. Therefore, activities towards 3-FBA could be determined directly from conversions at an earlier time-point (1 hour), using whole cells. For example, after 1 hour the conversion by D469T is <35%. Seven mutants selected randomly from the new library were analysed in detail for 3-FBA and showed that the specific activity in clarified lysates correlated well with those determined using whole cells in the high-throughput screen (see Supplementary Information Figure S2). At the highest specific activity reached, mass-transfer into whole cells appeared to have become limiting for 3-FBA.

For the 4-FBA and 3-HBA reactions, the yields at 18 hours determined from clarified lysates also correlated linearly with those from whole cells. Conversion yields after 18 hours using whole cells were approximately 50% of those using clarified lysates, suggesting a lower mass-transfer rate into the whole cells, for the aromatic substrates, compared to the HPA substrate (see Supplementary Information Figure S3). However, the good correlation of activities allowed us to robustly analyse and interpret the relative characteristics of all of the clones screened at high-throughput from each library on the three substrates, as shown in Figures 1 and 2.

3.1.1. Screening of theS385X/D469T/R520Q library

Saturation mutagenesis of S385 in wild-type *E. coli* TK was found previously to give no mutants with improved activity towards the linear substrates glycolaldehyde or propanal (Hibbert et al., 2007; Hibbert et al., 2008). By contrast, in the present study the whole-cell screening of the S385X/D469T/R520Q library gave at least five unique mutants with markedly improved (2-5.5 fold) final product yields from 3-HBA (relative to D469T), and two mutants with 2-2.5 fold improved yields from 4-FBA, where an improved yield indicates a higher activity as outlined above. Furthermore, one mutant with only modestly improved (1.8 fold) activity of 3-FBA (Figure 1) was obtained with whole cells. Most mutants actually lowered or maintained the conversion of 3-FBA after 1 hour with whole cells. Strikingly, the conversion yields from 3-HBA were highly correlated with those from 4-FBA, and to a lesser degree with the activities for 3-FBA. This suggested the same mechanism by which improvements are made, upon mutation of S385. Improvements for 3-FBA appears to arise through different mechanisms, including relief of substrate inhibition as discussed further below.

Mass transfer limitations in whole-cell screens potentially suppressed the observation of larger increases in activity (see Figure S2) and so specific activities in clarified lysates were also compared. With clarified lysates, D469T had an activity on 3-HBA of only 0.0071 µmol/mg/min, and a final product yield of only 7 % after 18 hours (Table 1). By contrast, the S385V/D469T/R520Q, S385Y/D469T/R520Q and S385T/D469T/R520Q mutants achieved 15%, 25%, and 31% final product yields, respectively. Their activities also increased proportionally, between 6- and 13-fold, to 0.03, 0.071 and 0.073 µmol/mg/min, respectively. The same trend was also observed for the conversion of 4-FBA but not with 3-FBA (Table 1). For 3-FBA, the only mutant with a higher (1.8-fold) activity in the high-throughput whole-cell screen was S385E/D469T/R520Q. Using clarified lysate, this mutant gave a 3.5-fold increase in specific activity with 3-FBA relative to that of D469T.

Most of the beneficial mutations (except S385E) introduced hydrophobic (A, V, Y) or polar (T) side chains, implicating modified hydrophobic interactions with each of the aromatic substrates, rather than affecting any specific interactions to the hydroxyl group of 3-HBA or the carboxylate moieties of 4-FBA or 3-FBA. This was slightly unexpected given that S385 interacts with the phosphate group in natural sugar substrates.

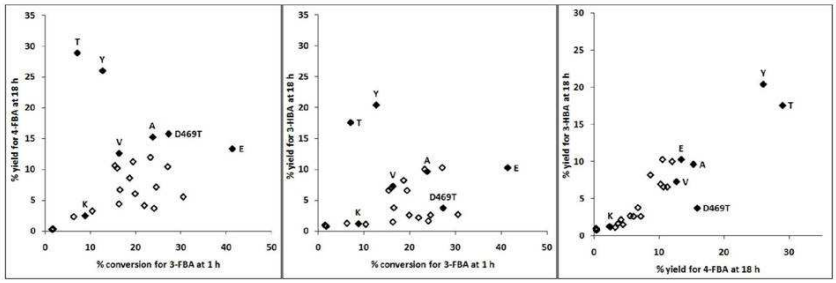


Figure 1. Pairwise comparisons of activities for the S385X/D469T/R520Q library. Activity on 3-FBA was measured directly from the conversion after 1 hour where the reactions were less than 35% complete. Final product yield at 18 hours gave an indirect measure of activity with 4-FBA and 3-HBA. All reactions were performed using freshly harvested whole cells.

3.1.2. Screening of the R358X/D469T/R520Q library

Saturation mutagenesis of R358 in wild-type *E. coli* TK previously led to only one beneficial mutant R358I, which gave only a 2-fold improvement in the activity towards the aliphatic aldehydes glycolaldehyde and propanal (Hibbert et al., 2007; Hibbert et al., 2008). Some triple mutants of R358X/D469T/R520Q, constructed and tested individually (data not shown), gave slightly reduced (R358H, and R358I) or severely diminished (R358E, R358Q) activities towards 3-FBA. The R358X/D469T/R520Q library provided no mutants with improved activity towards any of the three aromatic substrates (Figure 2). This is not surprising for 3-FBA and 4-FBA as we had previously identified charge stabilising interactions between their carboxylate moieties and R358. The initial high-throughput screen revealed that the conversions of 3-FBA and 4-FBA for all mutants were highly correlated, and that most gave less than 50% of the activity observed for D469T towards 3-FBA and 4-FBA. This is consistent with the stabilising electrostatic interactions between R358 and 3-FBA and 4-FBA. By contrast, half of the mutants retained an activity towards 3-HBA similar to that obtained with D469T. The conversions of 3-FBA and 4-FBA for all mutants were not well correlated at all to those of 3-HBA. This suggested that removing the positively charged R358 side-chain was not as detrimental for 3-HBA as it is for 3-FBA or 4-FBA. However, the lack of any improved R358 mutants towards 3-HBA was still surprising, as new favourable interactions with the uncharged hydroxyl moiety of 3-HBA might be expected.

A few mutants, R358T in particular, maintained at least 60% of the activity to 3-FBA and 4-FBA. By contrast, R358T significantly reduced the activity towards 3-HBA, which was particularly unexpected given the potential to form a new hydrogen bond to this substrate. Conversely, 3-HBA tolerated a histidine residue, whereas the negatively charged 4-FBA and 3-FBA did not. This indicates

that the aromatic substitution of 3-HBA is oriented differently towards R358 when bound within the active site, compared to 3-FBA and 4-FBA.

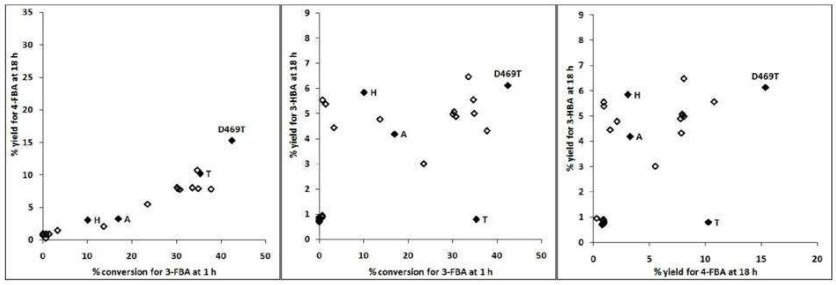


Figure 2. Pairwise comparisons of activities for the R358X/D469T/R520Q library. Activity on 3-FBA was measured directly from the conversion after 1 hour where the reactions were less than 35% complete. Final product yield at 18 hours gave an indirect measure of activity with 4-FBA and 3-HBA. All reactions were performed using freshly harvested whole cells.

Overall, the S385X library appeared to be more tolerant to different mutations than the R358X library when screening against 3-FBA and 4-FBA. However, the R358 residue was tolerant to most mutations with the activity on 3-HBA. It appeared then that the interactions between natural TK substrates and residue S385 were no longer critical or present with the aromatic substrates, whereas for 3-FBA and 4-FBA, the charge-charge interaction with R358 was still important as determined previously (Payongsri et al., 2012).

3.2. Kinetic parameters of the mutants with 3-FBA, 4-FBA, and 3-HBA

Any mutant with improved activity or yield towards at least one substrate was assessed in more detail with all three substrates to understand whether any shifts in specificity were due to the loss of affinity or increased catalytic efficiency (k_{cat}). All kinetic parameters are summarised in Table 1. S385E/D469T/R520Q gave a 3.5-fold increase in specific activity with 3-FBA relative to that with D469T, and it was found to significantly increase the k_{cat} to 34.3± 1.6 s^{-1}, which was 2.6 times greater than that for D469T. Furthermore, the K_M decreased to 18.3 ± 2.6 mM which was 3-fold lower than that for D469T, although most of this substrate binding effect was already achieved in with the R520Q mutation in D469T/R520Q. Therefore, most of the beneficial effect of the S385E mutation was due to an increased k_{cat}. The k_{cat}/K_M of S385E/D469T/R520Q was increased to 1900±270 s^{-1} M^{-1}, a 4-fold improvement over D469T/R520Q, and a 9-fold increase over D469T. Surprisingly, the substrate inhibition observed previously in all TK mutants with 3-FBA (Payongsri et al., 2012), was

also completely removed. No inhibition was observed up to the maximum concentration tested of 90 mM 3-FBA.

The specific activity of S385E/D469T/R520Q towards 3-HBA was also improved 6-fold relative to D469T, although a K_M of 245 mM and k_{cat} of 0.59 s^{-1} indicated that this activity was still relatively modest. The low activities made it too difficult to obtain kinetic parameters for D469T and D469T/R520Q. Interestingly, the same mutant lost 80% of the specific activity observed towards 4-FBA with D469T. In fact, the K_M (72 ± 30 mM) for 4-FBA improved 3.5-fold relative to that with D469T, but increased 3-fold relative to that with D469T/R520Q. Furthermore, the k_{cat} was 16-fold lower than for D469T, though marginally improved relative to D469T/R520Q.

It is interesting that S385E increases the k_{cat} for 3-FBA, but has no effect on the k_{cat} for 4-FBA. This suggests that k_{cat} for aromatic aldehyde substrates is dependent on the aromatic substitution type and position, and is not a general catalytic effect of the enzyme on all substrates. Indeed k_{cat} can be affected by substrate orientation and proximity to the cofactor and catalytic residues. The data suggest that while the carboxylate moieties of 3-FBA and 4-FBA both interact with R358, the reactive aldehyde moieties are oriented differently in the active site due to the different spatial constraints of the two molecules, giving rise to their respective k_{cat} values. S385 mutations appear to influence the molecular orientation and hence k_{cat}. Of course, electronic effects on the aldehyde moiety due to different positions of the aromatic substitutions in 3-FBA and 4-FBA may also potentially play a role, although 4-FBA would be expected to have greater reactivity than 3-FBA based on electronic effects alone.

The highest values of k_{cat}/K_M were found for S385Y/D469T/R520Q and S385T/D469T/R520Q towards 3-FBA (Table 1). Most of this improvement was due to an almost 10-fold decrease in K_M relative to D469T/R520Q. However, their activities towards 3-FBA were significantly lower than most of the other mutants, and only around 50% of D469T. This was due to significant inhibition of both mutants by 3-FBA at concentrations above 20 mM. This is in marked contrast to the removal of substrate inhibition by the S385E mutation. S385Y/D469T/R520Q and S385T/D46T/R520Q were also the only two mutants to give at least 40% greater conversion of 4-FBA during screening, than for D469T. Previously it was observed that the R520Q mutation in D469T/R520Q decreased the K_M 10-fold, but unfortunately also decreased the k_{cat} 25-fold with 4-FBA (Payongsri et al., 2012). The two triple mutants S385Y/D469T/R520Q and S385T/D469T/R520Q retained similar affinities towards 4-FBA as for D469T/R520Q, but their k_{cat} values were considerably improved. The restored k_{cat} and good substrate affinities of these two mutants resulted in a dramatic increase in their activities towards 4-FBA, a 4.5-5 fold improvement in k_{cat}/K_M relative to that of D469T, and over 10-fold

improvement in k_{cat}/K_M compared to D469T/R520Q. The latter is attributable to the effect of the S385Y and S385T mutations upon k_{cat} alone.

S385Y/D469T/R520Q and S385T/D469T/R520Q were also found to have greater activities towards 3-HBA compared to D469T. The k_{cat} and K_M values for 3-HBA could not be readily obtained for D469T or D469T/R520Q due to low activities. The k_{cat} values obtained for S385Y/D469T/R520Q and S385T/D469T/R520Q with 3-HBA were 2.1 and 1.0 s^{-1} respectively, which were similar to those for 4-FBA but 3 to 4 fold lower than those for 3-FBA. Furthermore, their K_M values were 100-300 times higher than those for 3-FBA, and 8-25 fold higher than for 4-FBA. Therefore, their k_{cat}/K_M values were up to 1000 times lower than for S385Y/D469T/R520Q with 3-FBA.

These data support our previous hypothesis that enzyme-substrate affinity and positioning of the substrate are competing major factors to be overcome when attempting to improve the bioconversion rate of poorly accepted aromatic aldehydes in transketolase (Payongsri et al., 2012). The specific activities of all mutants towards 3-HBA are significantly lower than with 3-FBA or 4-FBA, as their K_M values are all still much higher than the substrate concentration. However, different k_{cat} effects observed suggest positioning of the substrate relative to catalytic groups can have a significant impact. This underlies the different shifts in substrate preference for the S385E mutation within D469T/R520Q compared to those of S385Y and S385T. S385E mostly improved the k_{cat} towards 3-FBA, whereas for 4-FBA the greatest impact was an increase in K_M. S385Y and S385T both improved the K_M for 3-FBA, and yet increased the k_{cat} towards 4-FBA. This suggests that the position of the substrate relative to the catalytic groups is very sensitive to mutation, and that while 3-FBA and 4-FBA are both anchored to R358 via favourable electrostatic interactions, they orient the aldehyde moiety differently to the active-site cofactor. This simply reflects the different relative positions of their constrained aldehyde and carboxylate moieties.

	Substrate	Specific activity (relative to D469T)	Yield [b] (%)	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (s^{-1} M^{-1})
D469T	3-FBA	1.00[a]	67 (1.0)	56 (10)[c]	13.2 (1.5)	240 (50)
D469T/R520Q	3-FBA	0.75	67 (0.1)	13 (4)[c]	6.0 (0.75)	470 (170)
S385E/D469T/R520Q	3-FBA	3.5	63 (1.0)	18(3)	34 (1.6)	1900 (270)
S385Y/D469T/R520Q	3-FBA	0.47	53 (0.8)	1.3 (0.4)	7.0 (0.3)	5400 (1490)
S385T/D469T/R520Q	3-FBA	0.52	59 (0.5)	1.7 (0.25)	7.3 (0.2)	4250 (620)
D469T	4-FBA	1.00[a]	30 (1.7)	251 (240)[c]	5.0 (4.4)	20 (26)
D469T/R520Q	4-FBA	0.45	13 (0.4)	25 (8)[c]	0.20 (0.03)	8.1 (2.8)
S385E/D469T/R520Q	4-FBA	0.19	<5	72 (30)	0.3 (0.1)	4.6 (2.4)
S385Y/D469T/R520Q	4-FBA	3.38	48 (0.4)	15 (3)	1.7 (0.1)	110 (21)
S385T/D469T/R520Q	4-FBA	2.09	48 (1.7)	23 (6)	2.1 (0.3)	90 (26)
D469T	3-HBA	1.00[a]	7	n.d.[d]	n.d.	n.d.
D469T/R520Q	3-HBA	n.d.	n.d.	n.d.	n.d.	n.d.
S385E/D469T/R520Q	3-HBA	6.0	15 (1)	245 (15)	0.6 (0.1)	2.4 (0.2)
S385Y/D469T/R520Q	3-HBA	12.4	25 (1)	390 (10)	2.1 (0.2)	5.4 (0.1)
S385T/D469T/R520Q	3-HBA	12.7	31 (2)	180 (10)	1.0 (0.1)	5.5 (0.2)

Table 1.The kinetic parameters of all the mutants towards 3-FBA, 4-FBA and 3-HBA. [a] Specific activities of D469T at 50 mM 3-FBA/50 mM HPA, 30 mM 4-FBA/50 mM HPA and 15 mM 3-HBA/ 30 mM HPA were 4.63 µmol/mg/min, 0.45 µmol/mg/min, and 0.0071 µmol/mg/min, respectively. [b] Yields were obtained in all cases after 18 hours reaction. [c] The kinetic data of D469T and D469T/R520Q were from (Payongsri et al., 2012). [d] Values were not determined (n.d.) due to very low activity.

3.3. Computational modelling of S385E/Y/D469T/R520Q structure and substrate binding

The shifts in k_{cat} are not the same for all three substrates, suggesting that they are due to modified binding orientation or proximity to the catalytic residues, rather than general effects on catalytic enzyme residues or the cofactors. For 3-HBA, the proximity to R358 also appears to be less important than that of 3-FBA and 4-FBA. Computational modelling was used to rationalise these observations structurally.

3-FBA is expected to bind in the S385E/D469T/R520Q active-site in a rather different conformation than for D469T and D469T/R520Q. S385E/D469T/R520Q bound to 3-FBA was modelled in AUTODOCK and compared in Figure 3 to those obtained previously for D469T and

D469T/R520Q (Payongsri et al., 2012; Strafford et al., 2012). Docking of 3-FBA in S385E/D469T/R520Q revealed two possible binding clusters, with different binding energies and populations, as shown in Figure 3a. The first binding cluster had a lower binding energy, but also a lower population (38%) than the second. It resulted from interactions between R358 and H461, and the carboxylate moiety on 3-FBA, and was found in the previous models for D469T (Figure 3b) and D469T/R520Q (Figure 3c). However, D469T and D469T/R520Q previously gave two orientations within the cluster due to a 180° rotation of the aromatic ring. One orientation presented the aldehyde in a productive position at 4.89 Å (in D469T) from the cofactor-enamine intermediate, and the other at 7.21 Å. By contrast, the S385E/D469T/R520Q model only gave the least productive orientation. However, the more productive conformation is still plausible in S385E/D469T/R520Q, and an artefact of structure modelling or docking cannot be ruled out. The second cluster in S385E/D469T/R520Q was more populated (62%) than the first, and placed the aldehyde in amore catalytically productive location. Interestingly, the 3-FBA in this cluster was bound very differently, with no interaction between R358 and the carboxylate moiety of 3-FBA. Instead the carboxylate moiety formed hydrogen bonds with the side-chains of R91 and H26, and with the backbone NH of G25. The aldehyde was held in place by forming hydrogen bonds with H26 and H261, placing the aldehyde 4.63Å away from the reactive enamine intermediate. This distance is slightly shorter than the closest predicted previously with D469T (4.89Å) (Payongsri et al., 2012), and could contribute to the improved k_{cat}. The second cluster in S385E/D469T/R520Q also formed only one binding conformation across the entire population. This may improve k_{cat} entropically as less repositioning is required on average to move from the bound substrate conformations into a transition state for reaction. The new binding mode found only in S385E/D469T/R520Q could also explain the alleviation of substrate inhibition. It is possible that substrate inhibition is caused by the unproductive binding conformation within the first cluster, as it is common to all three structure models but significantly less populated in S385E/D469T/R520Q.

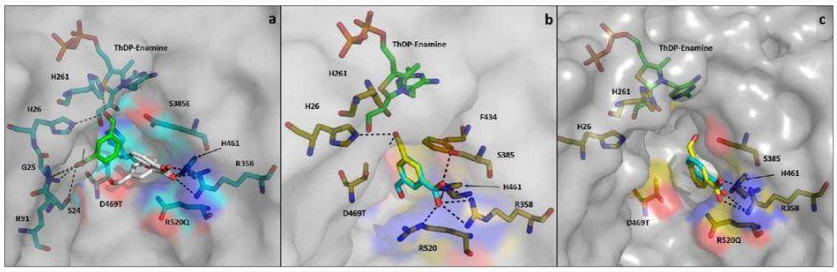


Figure 3. The computational docking of 3-FBA into a) S385E/D469T/R520Q, b) D469T, and c) D469T/R520Q active-sites. For S385E/D469T/R520Q the first cluster is shown with white sticks, and the second cluster is shown with green sticks. For D469T and D469T/R520Q, the two conformations within the single identified cluster are shown with yellow (catalytically productive) and cyan (catalytically unproductive) sticks. Calculated hydrogen bonds are shown as black dotted lines.

Computational docking was used to investigate the significantly improved k_{cat} found with the S385Y mutation in S385Y/D469T/R520Q for both 4-FBA and 3-HBA. Docking revealed only one binding cluster and conformation of 4-FBA in S385Y/D469T/R520Q (Figure 4a) which provides an entropic explanation for the improved k_{cat} relative to that of D469T/R520Q, for which many clusters were predicted (Payongsri et al., 2012). Autodock also previously predicted a single binding cluster for 4-FBA in D469T (Payongsri et al., 2012) which has a 3-fold higher k_{cat} than S385Y/D469T/R520Q (Table 1). The distance from the enamine intermediate to the aldehyde in 4-FBA was predicted to be 5.08 Å in S385Y/D469T/R520Q, which is within error of the 5.05 Å found in the closest conformer predicted with D469T (Payongsri et al., 2012). The predicted binding of 4-FBA in S385Y/D469T/R520Q also showed possible pi-stacking between 4-FBA and the tyrosine ring of S385Y. This pi-stacking, together with hydrogen bonding between the carboxylate moiety and R358, directed the binding of 4-FBA into a single conformation which could therefore improve the k_{cat} relative to that of D469T/R520Q. However, S385T provided similar gains in k_{cat}, suggesting pi stacking may not be the only important factor.

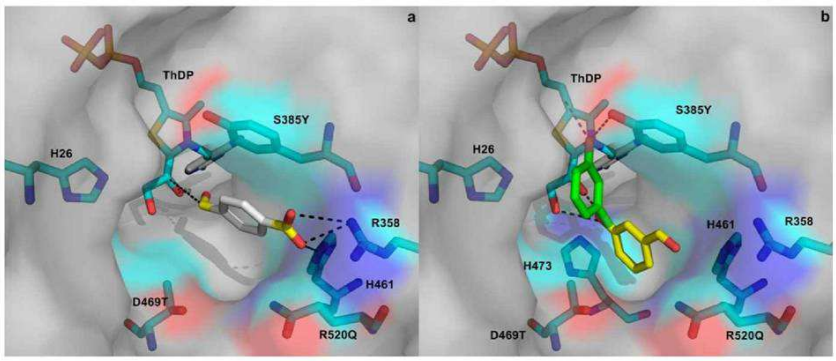


Figure 4. The computational docking of a) 4-FBA and b) 3-HBA into the S385Y/D469T/R520Q active site. For 4-FBA a single docking cluster was observed (white sticks). For 3-HBAtwo binding clusters were observed and shown in green sticks (catalytically productive) and yellow sticks (catalytically unproductive). Calculated hydrogen bonds are shown as black dotted lines.

Computational docking was also used to investigate the similarities and differences between 4-FBA and 3-HBA activity in S385Y/D469T/R520Q. This mutant improved the specific activity towards both 4-FBA and 3-HBA, but the activity towards 3-HBA was still much lower than for 4-FBA. The k_{cat} values for 4-FBA and 3-HBA were similar but their K_M values were at least 20-fold different.

Whereas 4-FBA could form pi stacking interactions with the tyrosine ring of S385Y, docking did not identify any binding cluster of 3-HBA with similar pi-stacking. In addition, it predicted only two out of the four major clusters to be deep within the active site. This prevalence of non-specific binding may in part explain the poor K_M for 3-HBA relative to that of 4-FBA.

The docking of 3-HBA in S385Y/D469T/R520Qsuggests that one of the clusters, highlighted in green in Figure 4b, would be the only catalytically productive cluster. The distance between the enamine intermediate and the aldehyde of 3-HBA in that cluster was 4.57 Å which is similar to that observed with 4-FBA. This at least would explain the similar k_{cat} values for 3-HBA and 4-FBA in S385Y/D469T/R520Q. Instead of the pi-stacking observed for 4-FBA, the hydroxyl moiety in 3-HBA was found to form a hydrogen bond with the S385Y tyrosine side-chain 4-hydroxyl group. 3-HBA was not found to interact with R358. This is consistent with the experimental data in which R358 mutations did not lead to improved activity towards 3-HBA, and also in particular why the R358T mutation was not beneficial.

Conclusion

Random mutation of one residue in the first shell of transketolase active site has shown to improve the enzyme activity and yield towards both carboxylated and hydroxylated aromatic aldehyde substrates. Triple mutants were created with up to 13-fold greater activity than D469T and the activities were good enough to determine the kinetic parameters of the enzymes towards aromatic hydrophobic aldehyde for the first time. Comparing the kinetic parameters between the carboxylated substrates and 3-HBA, the slow bioconversion of 3-HBA is likely to be due to low affinity between enzyme and substrate. Computational docking was able to rationalise the different behaviour of the three substrates in terms of their different response to mutations upon k_{cat}, K_M and substrate inhibition. The divergence of binding modes for the three aromatic substrates identified by kinetics and their structural interpretations, indicates that targeted saturation mutagenesis approaches are likely to require different libraries for further improving each of these activities.

Acknowledgement

The authors would like to thank the support from the Royal Thai government for Panwajee Payongsri and UCL Chemistry Department for David Steadman.

References

Asztalos, P., Parthier, C., Golbik, R., Kleinschmidt, M., Hubner, G., Weiss, M.S., Friedemann, R., Wille, G., Tittmann, K., (2007) Strain and near attack conformers in enzymic thiamin catalysis: X-ray crystallographic snapshots of bacterial transketolase in covalent complex with donor ketoses xylulose 5-phosphate and fructose 6-phosphate, and in noncovalent complex with acceptor aldose ribose 5-phosphate. Biochemistry 46, 12037-12052.

Bolte, J., Demuynck, C., Samaki, H., (1987) Utilization of enzymes in organic chemistry: Transketolase catalyzed synthesis of ketoses. Tetrahedron Lett. 28, 5525-5528.

Breuer, M., Hauer, B., (2003) Carbon–carbon coupling in biotransformation. Curr. Opin. Biotechnol. 14, 570-576.

Cazares, A., Galman, J.L., Crago, L.G., Smith, M.E., Strafford, J., Rios-Solis, L., Lye, G.J., Dalby, P.A., Hailes, H.C., (2010) Non-alpha-hydroxylated aldehydes with evolved transketolase enzymes. Org. Biomol. Chem. 8, 1301-1309.

Chen, B.H., Hibbert, E.G., Dalby, P.A., Woodley, J.M., (2008) A new approach to bioconversion reaction kinetic parameter identification. AIChE J 54, 2155-2163.

Dalmas, V., Demuynck, C., (1993) An efficient synthesis of sedoheptulose catalyzed by spinach transketolase. Tetrahedron Asymmetry 4, 1169-1172.

Demuynck, C., Bolte, J., Hecquet, L., Dalmas, V., (1991) Enzyme-catalyzed synthesis of carbohydrates: synthetic potential of transketolase. Tetrahedron Lett. 32, 5085-5088.

Enders, D., Narine, A.a., (2008) Lessons from nature: biomimetic organocatalytic carbon-carbon bond formations. J. Org. Chem. 73, 7857-7870.

Fessner, W.D., (1998) Enzyme mediated C-C bond formation. Curr Opin Chem Biol 2, 85-97.

Fessner, W.D., Helaine, V., (2001) Biocatalytic synthesis of hydroxylated natural products using aldolases and related enzymes. Curr. Opin. Biotechnol. 12, 574-586.

Galman, J.L., Steadman, D., Bacon, S., Morris, P., Smith, M.E., Ward, J.M., Dalby, P.A., Hailes, H.C., (2010) alpha,alpha'-Dihydroxyketone formation using aromatic and heteroaromatic aldehydes with evolved transketolase enzymes. Chem. Commun. (Camb.) 46, 7608-7610.

Galman, J.L., Steadman, D., Haigh, L.D., Hailes, H.C., (2012) Investigating the reaction mechanism and organocatalytic synthesis of alpha,alpha'-dihydroxy ketones. Org. Biomol. Chem. 10, 2621-2628.

Goodsell, D.S., Morris, G.M., Olson, A.J., (1996) Automated docking of flexible ligands: applications of AutoDock. J. Mol. Recognit. 9, 1-5.

Hailes, H.C., Dalby, P.A., Lye, G.J., Baganz, F., Micheletti, M., Szita, N., Ward, J.M., (2010) alpha,alpha '-Dihydroxy Ketones and 2-Amino-1,3-diols: Synthetic and Process Strategies Using Biocatalysts. Curr Org Chem 14, 1883-1893.

Hecquet, L., Bolte, J., Demuynck, C., (1996) Enzymatic synthesis of "natural-labeled" 6-deoxy-L-sorbose precursor of an important food flavor. Tetrahedron 52, 8223-8232.

Hibbert, E.G., Senussi, T., Costelloe, S.J., Lei, W., Smith, M.E., Ward, J.M., Hailes, H.C., Dalby, P.A., (2007) Directed evolution of transketolase activity on non-phosphorylated substrates. J. Biotechnol. 131, 425-432.

Hibbert, E.G., Senussi, T., Smith, M.E., Costelloe, S.J., Ward, J.M., Hailes, H.C., Dalby, P.A., (2008) Directed evolution of transketolase substrate specificity towards an aliphatic aldehyde. J. Biotechnol. 134, 240-245.

Humphrey, A.J., Parsons, S.F., Smith, M.E.B., Turner, N.J., (2000) Synthesis of a novel N-hydroxypyrrolidine using enzyme catalysed asymmetric carbon–carbon bond synthesis. Tetrahedron Lett. 41, 4481-4485.

Ingram, C.U., Bommer, M., Smith, M.E., Dalby, P.A., Ward, J.M., Hailes, H.C., Lye, G.J., (2007) One-pot synthesis of amino-alcohols using a de-novo transketolase and beta-alanine: pyruvate transaminase pathway in Escherichia coli. Biotechnol. Bioeng. 96, 559-569.

Kobori, Y., Myles, D.C., Whitesides, G.M., (1992) Substrate specificity and carbohydrate synthesis using transketolase. J. Org. Chem. 57, 5899-5907.

Martinez-Torres, R.J., Aucamp, J.P., George, R., Dalby, P.A., (2007) Structural stability of E. coli transketolase to urea denaturation. Enzyme Microb. Technol. 41, 653-662.

Morris, K.G., Smith, M.E.B., Turner, N.J., Lilly, M.D., Mitra, R.K., Woodley, J.M., (1996) Transketolase from Escherichia coli: A practical procedure for using the biocatalyst for asymmetric carbon-carbon bond synthesis. Tetrahedron Asymmetry 7, 2185-2188.

Myles, D.C., Andrulis, P.J., Whitesides, G.M., (1991) A transketolase-based synthesis of (+)-exo-brevicomin. Tetrahedron Lett. 32, 4835-4838.

Nilsson, U., Meshalkina, L., Lindqvist, Y., Schneider, G., (1997) Examination of substrate binding in thiamin diphosphate-dependent transketolase by protein crystallography and site-directed mutagenesis. J. Biol. Chem. 272, 1864-1869.

Paramesvaran, J., Hibbert, E.G., Russell, A.J., Dalby, P.A., (2009) Distributions of enzyme residues yielding mutants with improved substrate specificities from two different directed evolution strategies. Protein Eng. Des. Sel. 22, 401-411.

Patrick, W.M., Firth, A.E., Blackburn, J.M., (2003) User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. Protein Eng. 16, 451-457.

Payongsri, P., Steadman, D., Strafford, J., MacMurray, A., Hailes, H.C., Dalby, P.A., (2012) Rational substrate and enzyme engineering of transketolase for aromatics. Org. Biomol. Chem. 10, 9021-9029.

Resch, V., Schrittwieser, J.H., Siirola, E., Kroutil, W., (2011) Novel carbon-carbon bond formations for biocatalysis. Curr. Opin. Biotechnol. 22, 793-799.

Samland, A.K., Sprenger, G.A., (2006) Microbial aldolases as C-C bonding enzymes--unknown treasures and new developments. Appl. Microbiol. Biotechnol. 71, 253-264.

Schenk, G., Duggleby, R.G., Nixon, P.F., (1998) Properties and functions of the thiamin diphosphate dependent enzyme transketolase. Int. J. Biochem. Cell Biol. 30, 1297-1318.

Schneider, G., Lindqvist, Y., (1998) Crystallography and mutagenesis of transketolase: mechanistic implications for enzymatic thiamin catalysis. Biochim. Biophys. Acta 1385, 387-398.

Schörken, U., Sprenger, G.A., (1998) Thiamin-dependent enzymes as catalysts in chemoenzymatic syntheses. BBA - Protein Structure and Molecular Enzymology 1385, 229-243.

Smith, M.E.B., Hibbert, E.G., Jones, A.B., Dalby, P.A., Hailes, H.C., (2008) Enhancing and Reversing the Stereoselectivity of Escherichia coli Transketolasevia Single-Point Mutations. Adv. Synth. Catal. 350, 2631-2638.

Smith, M.E.B., Smithies, K., Senussi, T., Dalby, P.A., Hailes, H.C., (2006) The First Mimetic of the Transketolase Reaction. European J Org Chem 2006, 1121-1123.

Sprenger, G.A., Schorken, U., Sprenger, G., Sahm, H., (1995) Transketolase A of Escherichia coli K12. Purification and properties of the enzyme from recombinant strains. Eur. J. Biochem. 230, 525-532.

Strafford, J., Payongsri, P., Hibbert, E.G., Morris, P., Batth, S.S., Steadman, D., Smith, M.E., Ward, J.M., Hailes, H.C., Dalby, P.A., (2012) Directed evolution to re-adapt a co-evolved network within an enzyme. J. Biotechnol. 157, 237-245.

Takayama, S., McGarvey, G.J., Wong, C.-H., (1997) Enzymes in organic synthesis: recent developments in aldol reactions and glycosylations. Chem Soc Rev 26, 407-407.

Toone, E.J., Whitesides, G.M., (1991) Enzymes as Catalysts in Carbohydrate Synthesis. Enzymes in Carbohydrate Synthesis. American Chemical Society, pp. 1-22.

Wohlgemuth, R., (2009) C2-Ketol elongation by transketolase-catalyzed asymmetric synthesis. J. Mol. Catal., B Enzym. 61, 23-29.

Yi, D., Devamani, T., Abdoul-Zabar, J., Charmantray, F., Helaine, V., Hecquet, L., Fessner, W.D., (2012) A pH-based high-throughput assay for transketolase: fingerprinting of substrate tolerance and quantitative kinetics. Chembiochem 13, 2290-2300.

Ziegler, T., Straub, A., Effenberger, F., (1988) Enzyme-Catalyzed Synthesis of 1-Deoxymannojirimycin, 1-Deoxynojirimycin, and 1,4-Dideoxy-1,4-imino-D-arabinitol. Angew. Chem. Int. Ed. Engl. 27, 716-717.